Surface-Driven RNA Refolding by the OB-Fold Proteins of the *Trypanosoma brucei* Editosome

Vom Fachbereich Biologie der Technischen Universität Darmstadt

zur

Erlangung des akademischen Grades

eines *Doctor rerum naturalium*

genehmigte Dissertation von

Dipl. Biol. Christin Voigt

aus Darmstadt

1. Referent: Prof. Dr. H. Ulrich Göringer

2. Referent: Prof. Dr. Gerhard Thiel

Tag der Einreichung: 25.01.2017

Tag der mündlichen Prüfung: 05.04.2017

Darmstadt 2017

D 17

Die Arbeit wurde am Fachbereich Biologie der Technischen Universität Darmstadt in der Arbeitsgruppe Molekulare Genetik (Prof. Dr. H. Ulrich Göringer) angefertigt. Die Arbeit wurde durch den Sonderforschungsbereich 902: "Molecular Principles of RNA-based Regulation" der Deutschen Forschungsgemeinschaft (DFG) gefördert.

Teile der Arbeit sind in folgende Publikationen eingegangen:

Voigt, C., Dobrychłop, M., Kruse, E., Czerwoniec, A., Kasprzak, J.M., Bytner, P., Bujnicki, J.M., and Göringer, H.U. (2017). Surface-Driven RNA-Refolding by the OB-Fold Proteins of the *Trypanosoma brucei* Editosome. bioRxiv doi: 10.1101/099705. (submitted)

Leeder, W.M., Voigt, C., Brecht, M., and Göringer, H.U. (2016). The RNA Chaperone Activity of the *Trypanosoma brucei* Editosome Raises the Dynamic of Bound pre-mRNAs. Sci. Rep. 6, 19309.

Kruse, E., Voigt, C., Leeder, W.M., and Göringer, H.U. (2013). RNA Helicases Involved in U-Insertion /Deletion-Type RNA Editing. Biochim. Biophys. Acta. 1829, 835-41.

Summary

RNA editing in African trypanosomes is an essential mitochondrial RNA processing reaction. It is required to decode otherwise untranslatable primary transcripts. The reaction is characterized by the insertion and/or deletion of, in several transcripts, hundreds of uridylyl residues. It relies on a specific class of small, non-coding RNA molecules, known as guide (g)RNAs. Guide RNAs act as templates in the process. They base-pair to the highly structured pre- and partially edited mitochondrial mRNAs before a multi-protein complex, the 20S editosome, catalyzes the individual steps of the RNA editing cycle. The 20S editosome executes a chaperone-like RNA remodeling activity. The activity acts on the pre-edited transcripts by enhancing the structural dynamics of primarily U-nucleotides in order to "loosen" the structure of the pre-mRNA substrate molecules.

In **chapter II**, I present an *in vitro* assay, to experimentally monitor the consequences of the RNA chaperone activity for the entry of gRNAs. The assay uses short, gRNA-mimicking DNA oligonucleotides (gDNA) and probes the formation of gDNA-pre-mRNA hybrid molecules by ribonuclease H (RNaseH) digestion. The data demonstrate that the RNA chaperone activity stimulates the formation of gDNA/pre-mRNA hybrids thereby acting as an early facilitator of the editing reaction.

The molecular nature of the editosome-inherent RNA chaperone activity is unknown. In **chapter III**, I test the hypothesis, that the oligonucleotide/oligosaccharide binding (OB)-fold proteins of the editosome are responsible for the activity. Importantly, these proteins are predicted to be in large parts intrinsically disordered. Furthermore, experimental evidence suggests that the proteins form a, structurally important, clustered subdomain within the editosome. To examine their potential chaperone activity I expressed the different proteins as recombinant polypeptides, both, as full-length and as OB-fold-only constructs. I verified the disorder prediction by circular dichroism (CD)-spectroscopy and analyzed their homo- and hetero-oligomerization behavior by size exclusion chromatography. Importantly, all OB-fold protein constructs show RNA chaperone activity with half maximal activities in the nM concentration range. The measured activities directly correlate to the surface areas of the different protein complexes suggesting that the RNA remodeling process is mainly surface-driven.
The analysis of the induced RNA structure changes was performed with nucleotide resolution using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE). The experiments demonstrate that the effect of the homo-oligomeric OB-fold protein complexes is qualitatively and quantitatively similar to the effect of 20S editosomes. The data further suggest an involvement of the intrinsically disordered protein regions. A coarse grained structural model of the 20S editosome is presented, which supports a preferential positioning of disordered protein domains on the surface of the complex and of ordered protein regions in the core of the editosome.

An analysis, to address possible RNA substrate and nucleotide specificities of the OB-fold protein-mediated RNA chaperone activity, is presented in **chapter IV**. For that RNaseH-based gDNA annealing assays were performed comparing the OB-fold of the editosome protein TbMP24 with 20S editosomes. Two different mitochondrial transcripts were used in the analysis: the pre-mRNA encoding ribosomal protein S12 (RPS12) and the pre-mRNA for apocytochrome b (CYb). I analyzed the RPS12 transcript at ten different regions of different structure and sequence contexts. The experiments demonstrate that both, TbMP24-OB and 20S editosomes, enhance the accessibility of a pre-mRNA for gDNA entry at every position in the RNA. The TbMP24-OB activity shows no preference for distinct RNA structures, but a preference for U-rich sequence regions. This U-preference was confirmed by SHAPE-experiments of the RPS12-RNA with TbMP24-OB.
Both TbMP24-OB and 20S editosomes act on both transcripts, with a slightly higher activity on the larger CYb-RNA. TbMP24-OB thereby shows the same overall substrate preference as 20S editosomes, which enables the protein to function as a model protein for studying the RNA chaperone activity of the 20S editosome.

RNA editing is differentially regulated throughout the lifecycle of *Trypanosoma brucei*, but the cause of that differential editing is unknown. The intracellular location of RNA editing, the mitochondrion, differs in the lifecycle stages, ultimately generating two different redox environments. In **chapter V**, I test whether the editosomal OB-fold proteins can serve as "redox switches". I uncovered that some of the OB-fold proteins of the editosome are sensitive towards changes in the redox environment, leading to altered thermal stabilities and altered oligomerization characteristics. To investigate the effects of oxidation on protein oligomerization, I mutated the cysteine codons in the TbMP24-OB-coding gene. The exchange of all cysteines by alanines resulted in a mutated protein with an altered oligomerization behavior. The mutations cause an inability to switch the oligomerization state, resulting in a protein variant that stays dimeric, comparable to the wild-type protein in its reduced form. The interaction with the binding partner TbMP18 is not hindered by the cysteine mutations, but the ratio of interacting proteins is altered. These findings have implications for the assembly of the editing complex *in vivo*, suggesting that cysteine thiol redox switching could be a mechanism to regulate editing in a lifecycle-dependent manner.

Zusammenfassung

RNA-*Editing* in afrikanischen Trypanosomen ist eine essenzielle RNA-Prozessierungsreaktion. Sie wird benötigt, um mitochondriale Primärtranskripte, die sonst nicht translatierbar wären, zu entschlüsseln. Die RNA-*Editing*-Reaktion ist durch das Einfügen und/oder Entfernen von, in einigen Fällen hunderten, Uridylatresten gekennzeichnet. Sie ist auf eine spezifische Klasse von kleinen, nicht-kodierenden RNA-Molekülen angewiesen, die *guide* (g)RNAs. Die gRNAs stellen die Vorlage für diesen Prozess dar. Sie bilden Basenpaare mit den hochstrukturierten mitochondrialen prä- und partiell edierten mRNAs, bevor ein Multiproteinkomplex, das 20S Editosom, die einzelnen Schritte des RNA-*Editing*-Zyklus katalysiert. Das 20S Editosom führt eine chaperonartige RNA-Strukturumformung durch. Diese Aktivität wirkt auf prä-edierte Transkripte, wobei die strukturelle Dynamik, primär von U-Nukleotiden, erhöht wird, um die Struktur dieser Substrat-RNAs "aufzulockern".

In **Kapitel II** präsentiere ich ein *in vitro* Testsystem, um die Konsequenzen der RNA-Chaperonaktivität auf die Bildung des prä-mRNA-gRNA-Hybrids zu untersuchen. In diesem Testsystem verwende ich kurze, gRNA nachahmende, DNA-Oligonukleotide (gDNA) und weise die Bildung von gDNA/prä-mRNA-Hybriden durch den Verdau mittels Ribonuklease H (RNaseH) nach. Die Daten zeigen, dass die RNA-Chaperonaktivität die Bildung von gDNA/prä-mRNA-Hybriden stimuliert. Damit ermöglicht die RNA-Chaperonaktivität die darauffolgende RNA-*Editing*-Reaktion.

Die molekulare Ursache der Editosom-inhärenten RNA-Chaperonaktivität ist unbekannt. In **Kapitel III** überprüfe ich die Hypothese, dass die Proteine des Editosoms, die ein Oligonukleotid/Oligosaccharid Bindemotif (*OB-fold*) besitzen, verantwortlich für diese Aktivität sind. Vorhersagen deuten darauf hin, dass diese Proteine in großen Teilen intrinsisch unstrukturiert sind. Weiterhin belegen Experimente, dass diese Proteine eine strukturell wichtige Subdomäne innerhalb des Editosoms bilden. Um die RNA-Chaperonaktivität der Proteine zu untersuchen, habe ich sie als rekombinante Polypeptide hergestellt, sowohl als Konstrukt in voller Länge, als auch als Konstrukt welches nur aus dem *OB-fold* besteht. Ich bestätigte die Vorhersage der unstrukturierten Proteinbereiche mittels Circulardichroismus-Spektroskopie und analysierte das Homo- und Hetero-Oligomerisierungsverhalten durch Größenausschluss-Chromatographie. Alle *OB-fold*-Proteinkonstrukte zeigen eine RNA-Chaperonaktivität mit halbmaximalen Aktivitäten im nM Konzentrationsbereich. Die gemessene Aktivität korreliert direkt mit der Oberfläche der unterschiedlichen Proteinkomplexe, was darauf hindeutet, dass der Prozess hauptsächlich durch die Proteinoberflächen bedingt ist.
Die Analyse der induzierten RNA-Strukturänderungen wurde mit Nukleotidauflösung mittels *Selective 2'-Hydroxyl Acylation analyzed by Primer Extension* (*SHAPE*) durchgeführt. Die Experimente zeigen, dass der Effekt der homooligomeren *OB-fold* Proteinkomplexe qualitativ und quantitativ dem Effekt von 20S Editosomen entspricht. Des Weiteren deuten die Daten darauf hin, dass intrinsisch unstrukturierte Proteinregionen involviert sind. Ein grobgranulares Strukturmodell des 20S Editosoms wird präsentiert, das die präferenzielle Anordnung dieser Bereiche an der Oberfläche des Komplexes und die Anordnung von strukturierten Bereichen im Kern des Komplexes unterstützt.

Eine Analyse, die mögliche RNA-Substrat- und Nukleotidpräferenzen der *OB-fold* vermittelten RNA-Chaperonaktivität betrachtet, wird in **Kapitel IV** gezeigt. Dafür wurden Experimente mit dem RNaseH-basierten gDNA Anlagerungstestsystem durchgeführt, die den *OB-fold* des editosomalen Proteins TbMP24 und 20S Editosomen vergleichen. Zwei unterschiedliche mitochondriale Transkripte wurden in dieser Analyse verwendet: die prä-mRNA, die für das ribosomale Protein S12 (RPS12) kodiert und die prä-mRNA für Apocytochrom B (CYb). Ich analysierte das RPS12 Transkript an zehn verschiedenen Positionen, die unterschiedliche Sequenz- und Strukturmerkmale aufweisen. Die Experimente zeigen, dass sowohl TbMP24-OB als auch 20S Editosomen die Zugänglichkeit der prä-mRNA für die Anlagerung der gDNA an jeder Position der RNA erhöhen. Die TbMP24-OB-Aktivität zeigt keine Präferenz für bestimmte Strukturmerkmale, jedoch eine Präferenz für U-reiche Regionen. Die U-Präferenz wurde mittels *SHAPE*-Experimenten, der RPS12-RNA mit TbMP24-OB, bestätigt.
Sowohl TbMP24-OB als auch 20S Editosomen zeigen eine Aktivität auf beiden Transkripten, mit einer leicht erhöhten Aktivität auf der größeren CYb-RNA. Somit hat TbMP24-OB die gleichen Substratpräferenzen wie das 20S Editosom. Damit kann es als Modellprotein verwendet werden, um die RNA-Chaperonaktivität des 20S Editosoms zu untersuchen.

RNA-*Editing* ist während des Lebenszyklus von *Trypanosoma. brucei* differenziell reguliert, jedoch ist die Ursache dafür unbekannt. Der intrazelluläre Ort an dem RNA-*Editing* stattfindet, das Mitochondrion, unterscheidet sich während der unterschiedlichen Stadien des Lebenszyklus, was zu zwei verschiedenen Redox-Umgebungen führt. In **Kapitel V** untersuche ich, ob die *OB-fold* Proteine des Editosoms, als "Redox-Schalter" dienen können. Ich zeige, dass einige der *OB-fold* Proteine des Editosoms sensitiv gegenüber einer Veränderung der Redox-Umgebung sind, was zu einer veränderten thermischen Stabilität und zu einem veränderten Oligomerisierungsverhalten führt. Um die Effekte der Oxidation auf die Proteinoligomerisierung zu untersuchen, mutierte ich die Cystein-Codons des TbMP24-OB-kodierenden Gens. Der Austausch aller Cysteine durch Alanine, führt zu einem mutierten Protein, das ein verändertes Oligomerisierungsverhalten aufweist. Die Mutationen führen dazu, dass die Proteinvariante ihren Oligomerisierungszustand nicht mehr ändert und als Dimer,

vergleichbar mit dem Wildtypprotein in seiner reduzierten Form, bestehen bleibt. Die Interaktion mit dem Bindepartner TbMP18 ist durch die Cysteinmutationen nicht beeinträchtigt, jedoch ist das Verhältnis der beiden interagierenden Proteine im Komplex verändert. Diese Erkenntnisse sind von Bedeutung für die Assemblierung des *Editing*-Komplexes *in vivo*, was darauf hindeutet, dass Redox-Schalter in die Regulation des RNA-*Editings*, in Abhängigkeit vom Lebenszyklus, involviert sein können.

Chapter I

General Introduction

The paragraph “RNA Helicases Involved in U-Insertion/Deletion-Type RNA Editing” has been included into:

Kruse, E., Voigt, C., Leeder, W.M., and Göringer, H.U. (2013). RNA Helicases Involved in U-Insertion /Deletion-Type RNA Editing. Biochim. Biophys. Acta. 1829, 835-841.

RNA Structure

Ribonucleic acid (RNA) was first discovered as "messenger" molecule, which transmits information between the information storage, DNA, and the catalytically active proteins (Crick, 1958). The discovery that the protein synthesis machinery largely relies on non-informational RNAs – with ribosomes in large parts being composed of RNA and the adapter molecule, transfer RNA, being completely constituted of RNA – led soon to the hypothesis of an RNA-based origin of life (Orgel, 1968, Crick, 1968). However, only the discovery of two enzymatic active RNAs, a self-splicing exon and ribonuclease P (Kruger et al., 1982, Guerrier-Takada et al., 1983), prompted the RNA world hypothesis (Gilbert, 1986). It suggests a stage in the early development of life, where protein and DNA were absent and self-replicating RNAs served as both information memory and functional molecule. Whether this hypothesis is true or false, still in our actual DNA, RNA and protein world, RNA is of great importance. RNAs are involved in multiple processes, for example small nuclear (sn)RNAs involved in splicing or micro (mi)RNA involved in regulation of translation. But we are just at the beginning of understanding the complexity of RNA-dependent processes, as exemplified by long non-coding (lnc)RNAs that are autonomously transcribed non-protein coding RNAs, with about 10000 different ones in humans. The lncRNAs are involved in a diversity of processes, *e.g.* the transcriptional silencing of one X-chromosome or the regulation of DNA methylation (reviewed by Cech and Steitz, 2014). Recently, several studies suggested a tight interaction network between the different kinds of RNAs, which leads to an even higher degree of complexity (Lu et al., 2016).

What determines the function of RNA? There are different layers of information in an RNA molecule, as the sequence of the four different nucleobases carries the genetic information in the RNA, but is also the determinant of RNA structure. This structure influences or determines the function of every RNA molecule. To derive the RNA structure from the sequence, a complex RNA folding problem has to be solved (Draper, 1992, Herschlag, 1995). There are three properties that determine the folding of RNA: (1) There are six rotatable dihedral angles in the phosphate-ribose backbone for each nucleotide, resulting in a huge amount of possible different folds. (2) There is a dominance of only a few base pairing interactions (Watson-Crick-, wobble-base pairing) that results in many nearly degenerate secondary structures. (3) Secondary structures are stabilized cooperatively by a high number of base pairing and stacking interactions. These properties result in a great complexity of the RNA folding problem, leading to many possible different folds excluding each other. The complexity is reduced by the hierarchical folding of the RNA. First, secondary structure is formed, by hydrogen bonding between the four different nucleobases, whereby base stacking interactions stabilize the duplexes. Second the secondary structural elements are spatially arranged to yield the three-dimensional tertiary structure. This hierarchical folding on the other hand gives rise to large energetic barriers, resulting in a rugged free energy landscape with deep local energy minima (Figure I.1A). These particular characteristics of RNA-folding result in long-lived structural variants, so called static heterogeneity (Schröder et al., 2004, Marek et al., 2011), which was intensively investigated by single molecule spectroscopy (Alemán et al., 2008). Different structures of RNAs with the same primary sequence can either have different functions, as it was shown for a synthetic ribozyme (Schultes and Bartel, 2000), the same function, as in the hairpin ribozyme from tobacco ringspot virus satellite RNA (Ditzler et al. 2008), or can be non-functional and thereby misfolded. In the cell, there are proteins that alter RNA folding and help to overcome energetic barriers.

RNA Structure Remodeling Proteins

As RNAs rely on a distinct structure to fulfill their tasks, correct folding is essential. There are different proteins that can remodel RNA to reach the desired fold (Figure I.1B). These can act either by specifically binding to the correctly folded RNA and thereby acting as stabilizer, by taking the desired fold out of the folding equilibrium, or via RNA remodeling by shifting the dynamic equilibrium. The different protein groups performing RNA structure remodeling are reviewed by Rajkowitsch et al., (2007) and introduced in the following sections.

RNA Annealers

RNA annealing is energetically favored and thereby proceeds *in vitro* without any protein action, although proteins with RNA annealing activity can accelerate the reaction. These proteins act by binding to one or two RNA strands to promote duplex formation. That can be achieved by a molecular crowding mechanism, where the local concentration of RNA is increased by binding to two strands simultaneously (Rajkowitsch et al., 2007). Such a mechanism was shown for the bacterial protein Hfq, which has separate binding surfaces for a small non-coding RNA and mRNA, comprising different nucleotide specificities (Panja and Woodson, 2012). A special case of RNA annealers are proteins with a matchmaking activity. They alter the structure of the bound RNA to expose the nucleobases, thereby transferring it to a configuration of higher annealing competence. These proteins are involved in processes, where a guide (g)RNA has to bind base-complementary to a larger target sequence, for example in ribosomal (r)RNA processing where the Imp3p and Imp4p proteins facilitate binding of the U3 small nucleolar RNA to the pre-rRNA (Gerczei and Correll, 2004). Such proteins with RNA annealing activity stay bound to a single stranded RNA until the duplex is faithfully formed.

RNA Helicases

RNA helicases are proteins that utilize ATP as an external energy source for the active unwinding of structured RNAs. The global activity of RNA helicases in the cell was shown by transcriptome-wide probing *in vivo* (Rouskin et al., 2014). RNA helicases are closely related to DNA helicases, with the eukaryotic ones belonging to the helicase super-families 1 and 2. All included families share a common helicase core of two highly similar RecA-like helicase domains. Inside a family there are characteristic conserved sequence motifs, like the DEAD-box motif. RNA helicases act by utilizing two different mechanisms. Most DNA helicases and viral RNA helicases perform canonical duplex unwinding, by binding to a single stranded region adjacent to the duplex and moving with defined directionality along the duplex, whereby the complementary strand is replaced. The translocation is dependent on the hydrolysis of ATP. The second mechanism, termed local strand separation, is used by DEAD-box helicases. Thereby the helicase binds directly to the duplex, which is assisted by adjacent single-stranded regions. It opens the duplex locally with an activity that has no directionality. The binding of an ATP molecule is a prerequisite for the binding of an RNA, while its hydrolysis is needed for the efficient release of the RNA. No clear RNA sequence or structure preferences were found *in vitro* although DEAD-box RNA helicases are involved in unique cellular processes. There are accessory segments flanking the common helicase domain, which were recently found to be highly disordered in a group of human DEAD-box helicases. These domains are discussed to be involved in protein-protein interactions modulating the specificity in the cellular context (reviewed by Leitão et al., 2015).

RNA Chaperones

RNA chaperones are proteins that change RNA structure independently of an external energy source. They prevent RNA folding from being trapped in local kinetic minima or help to overcome such minima. The term RNA chaperone is used for proteins, for which an RNA folding activity has been verified *in vivo* on its natural target, otherwise its activity is called RNA chaperone activity. Proteins with RNA chaperone activity are an inhomogeneous group. They do not share structure or sequence features and are multifunctional. They are characterized by an ATP-independent structure opening and are dispensable once the RNA has folded (Semrad, 2011). There is indication that tight RNA binding is detrimental to RNA chaperone activity (Mayer et al., 2007, Herschlag et al., 1994). This can be explained by the mechanism that is typically used by RNA chaperones. RNA chaperones act by binding single stranded RNA more tightly than double stranded RNA, thereby stabilizing the less folded transition states and destabilizing folding intermediates (Figure I.1C). The RNA chaperone has to dissociate, so that the RNA can refold. Therefore, the RNA chaperone complex has to be less stable than the natively folded RNA (Woodson, 2010).

How can these proteins unwind RNA independently of ATP? Within this context, intrinsic disordered protein regions gained attention. RNA chaperones are the protein group with the highest amount of intrinsic protein disorder (average of 54%) and the activity is linked almost exclusively to the disordered regions. The usage of intrinsically disordered regions for binding enables binding promiscuity, which is important because RNA chaperones have to act on a wide variety of misfolded substrates. The disorder offers an explanation for the mechanism of unwinding. One can view the disorder as an entropy source covering the energy demand of unfolding by structuring of the disordered chaperone when it binds to the RNA. This mode of action, where entropy is transferred from a protein to an RNA, is called the entropy transfer model of chaperone action (Tompa and Csermely, 2004).

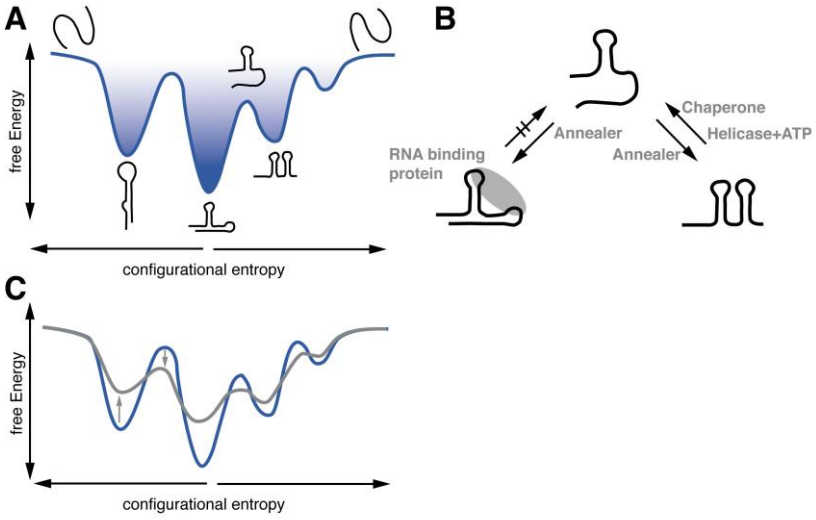


Figure I.1. RNA Folding and RNA Remodeling Proteins.
A. Sketch of a simplified two-dimensional RNA folding landscape. For clarity, only secondary structures are shown.
B. RNA remodeling proteins influencing RNA folding by shifting the folding equilibrium.
C. RNA chaperones smooth the RNA folding landscape by destabilizing intermediates and stabilizing transition states. The blue curve shows a folding landscape without and the gray curve with an RNA chaperone. Gray arrows indicate the changes.

RNA Editing in African Trypanosomes

African trypanosomes are unicellular, eukaryotic flagellates. Subspecies of *Trypanosoma brucei* cause diseases like the human African trypanosomiasis. In their lifecycle the parasitic trypanosomes undergo a host shift between an insect and a mammalian host requiring major adaptations of the parasite, resulting in different life cycle forms (see chapter V). The most important forms – the procyclic form that lives in the midgut of the tsetse fly and the mammalian blood-stream form – can be cultured *in vitro*. *T. brucei* belongs to the group of kinetoplastida, which is named after the presence of a kinetoplast, which is a structure located to the attachment site of their flagella. It is composed of circular DNAs, called minicircles and maxicircles, comprising the genome of the single mitochondrion. These circular DNAs are tightly connected to form a network that ensures that the complete genome is segregated and distributed in both daughter mitochondria (Jensen and Englund, 2012).

The maxicircles encode mitochondrial proteins of respiratory chain complexes, the F_0F_1-ATPase, the ribosome and ribosomal RNAs and are thereby essential for the organism. Some of the genes, encoding pre-mRNAs for these proteins, are heavily encrypted and have to be decrypted post-transcriptionally by the insertion and/or deletion of uridylyl residues. This unique type of RNA editing is catalyzed by a multi-protein complex and guiding (g)RNAs are used to determine the sites of editing. The gRNAs are encoded almost exclusively by the minicircles and are utilized as templates for the RNA editing reaction. This results in extreme cases in transcripts where fifty percent of the information is generated by RNA editing, including the generation of start and stop codons and the open reading frame. Such transcripts, edited over their full length, are called pan-edited RNAs. Other transcripts get only marginally edited and some have not to be edited at all to generate functional mRNAs (Göringer, 2012). The RNA editing is differentially regulated during the life cycle of the parasite. Until now the cause of this regulation is not known. Differential RNA editing is further examined in chapter V.

The RNA editing reaction is catalyzed by a 0.8 megadalton sized multi-protein complex with an apparent Svedberg (S) value of 20S and is thereby called the 20S editosome. The protein complex carries all the needed enzymatic activities to fulfill a full round of RNA editing (Figure I.2C, F) (reviewed by Göringer, 2012). The gRNA interacts with the pre-mRNA by forming a three-helix junction (Figure I.2B). The first editing site is defined by a non-base-paired nucleotide in 5'-direction of the so-called anchor duplex. The editing cycle is initiated by the first enzymatic step, when one of three endonucleases cleaves the pre-mRNA at the editing site. In deletion-type editing one of the two available exonucleases cleaves off U nucleotides at the 3'-end of the 5'-cleavage product, whereas a terminal uridylyl transferase (TUTase) adds Us in insertion type RNA editing. Two ligases are present that join the 3'- and 5'-end after RNA editing took place. Several editing cycles have to be performed until one editing block, defined by one gRNA, is processed. Multiple gRNAs are needed for the complete processing of a pre-mRNA. The RNA editing proceeds with a general 3'- to 5'-directionality on the pre-mRNA.

A second editosomal complex, the 35-40S editosome, resembles the complex that is actively engaged in the RNA editing reaction. It likely consists of the 20S editosomal proteins, pre-mRNAs and gRNAs. Whereas the 20S editosome has a nanomolar affinity for gRNA, pre-mRNA and hybrids of both RNAs, the 35-40S editosome does not show binding or RNA editing activity *in vitro*, presumably because the binding site is occupied by endogenous RNAs. The structures of the 20S and 35-40S editosomes have been analyzed by cryo-electron microscopy (cryo-EM), resulting in averaged coarse grained structures of the two complexes (Golas et al., 2009) (Figure I.2D, E). The 20S editosome consists of two globular subdomains, while the larger 35-40S editosome shows a flat platform on one side and a semi-spherical back. Comparative analysis of the two structures showed that the two complexes share a structural core and the 20S editosome may comprise the major part of the platform of the 35-40S editosome, whereas the semi-spherical back is variable. Analysis of the variability of different 35-40S editosome populations revealed also that the semispherical domain is the major source of variability, indicative of the region being composed of RNA that is variable in size. The molecular mass of the 20S editosome suggests that each protein is present in a single copy, although the arrangement of the proteins is so far unknown (Golas et al., 2009).

Partial structures of five editosomal protein components and protein dimers were solved by crystallization and X-ray diffraction (reviewed by Czerwoniec et al., 2015). Interactions between the different proteins were probed by yeast-two-hybrid experiments, co-expression in *E. coli* (Schnaufer et al., 2010) and recently by cross-linking of the proteins (McDermott et al., 2016). The studies suggest a structure of two subcomplexes for insertion and deletion RNA editing that are tightly inter-connected, with a set of proteins containing an oligonucleotide /oligosaccharide binding fold (OB-fold) being involved in this interconnection (see chapter III). The finding that editosomes have only one mRNA binding site (Böhm et al., 2012) supports the tight inter-connection of the insertion and deletion sub-complexes, forming a single active center. Although there is already some knowledge about protein structures and interactions, the arrangement of the proteins in the complex is so far not solved.

Structure of RNAs Involved in U-Insertion/Deletion-Type RNA Editing

mRNA

The pre-mRNAs are encoded by the mitochondrial maxicircle genome that is transcribed to multicistronic RNAs. Neither the promoters of the RNA polymerase nor the processes that lead to the generation of

monocistronic transcripts are known so far. The ends of the pre-mRNA consist of short untranslated regions (UTRs) with a monophosphate group at the 5' end (Aphasizhev and Aphasizheva, 2011). The pre mRNAs have a 3' poly(A) tail of about 20nt that is added by polyadenylation with the kinetoplast poly(A) polymerase (KPAP1) (Etheridge et al., 2008). This polyadenylation has an opposing effect on pre-edited and edited transcripts. It destabilizes pre-edited transcripts while stabilizing partially and fully edited RNAs after a few editing events took place (Kao and Read, 2005). The short poly(A) tail is extended by the addition of a poly(A/U) tail (120-250nt) after the transcript became fully edited, involving KPAP1, the terminal uridylyl transferase RET1 and two pentatricopeptide repeat proteins (KPAF1, KPAF2) (Etheridge et al., 2008, Aphasizheva et al., 2011). Never edited transcripts are also processed by the addition of a long poly(A/U) tail. This tail is important, because only mRNAs with this 3' extension are translated by the ribosome (Aphasizheva et al., 2011). Regarding the coding region, especially pre-edited transcripts that have to get pan-edited have an extreme nucleotide bias towards purins. This bias results in G-clusters, forming up to five G-quadruplex (GQ) structures in one transcript. GQs consist of multiple stacked arrays of each four G-residues hydrogen bonded via the Hoogsteen edge (Leeder et al., 2016b). GQs are thermodynamically highly stable and were found to suppress translation (Endoh et al., 2013). RNA editing resolves the GQs by insertion of uridylyl residues. As a consequence of RNA editing the mRNAs comprise an U-bias, with more than fifty percent Us in pan-edited transcripts. Never edited and marginally edited transcripts have a bias similar to edited transcripts with about fifty percent Us (Leeder et al., 2016b). Chemical probing by using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) showed that, *in vitro*, the pre-mRNAs and never edited RNAs are highly structured, comprising aside from the mentioned GQ-elements almost all known secondary structural elements (Leeder et al., 2016a). Thus the substrate for RNA editing is a highly structured pre-mRNA.

gRNA

gRNAs have a length of 50-70nt, with an oligo(U) tail of about 15Us at their 3'- end and a 5'-triphosphate (Blum and Simpson, 1990). Most of the gRNAs are encoded by the minicircles with genes flanked by imperfect inverted repeats. The transcription starts about 30bp apart from the upstream repeat in a conserved sequence (Pollard et al., 1990). Recently a model was proposed, where long precursors of the gRNAs are transcribed from the sense and the anti-sense strand of minicircles, generating transcripts that are partially base complementary. The maturation of the gRNA is thought to be performed in a protein complex, called the mitochondrial 3'-processome. It involves poly-uridylation of the primary transcript followed by 3'- to 5'-exonuleolytic digestion, until the duplex region of sense and anti-sense transcript is reached. Afterwards both 3' ends get uridylated again and the antisense gRNA gets degraded (Suematsu et al., 2016). The structure of the mature gRNAs is of low thermodynamic stability with T_ms of 33-39°C. The structure was shown to be a double stem-loop structure, with a less thermodynamically stable 5'-stem (stem I) and a more stable 3'-stem (stem II) (Figure I.2A). The structure is conserved among different gRNAs. On the one hand it enables an easy unfolding of the stem I to facilitate gRNA-pre-mRNA-hybrid formation and on the other hand it provides distinct structural features for the recognition by proteins (Schmid et al., 1995). The detailed analysis of one gRNA revealed a 3D working model, where the two stems are tightly packed in a parallel configuration with their major grooves facing each other and a flexible U-tail. The surface of the molecule is in big parts formed by the sugar-phosphate backbone (Hermann et al., 1997). The model was later supported by a crystal structure of a gRNA bound to a gRNA binding protein complex (Schumacher et al., 2006).

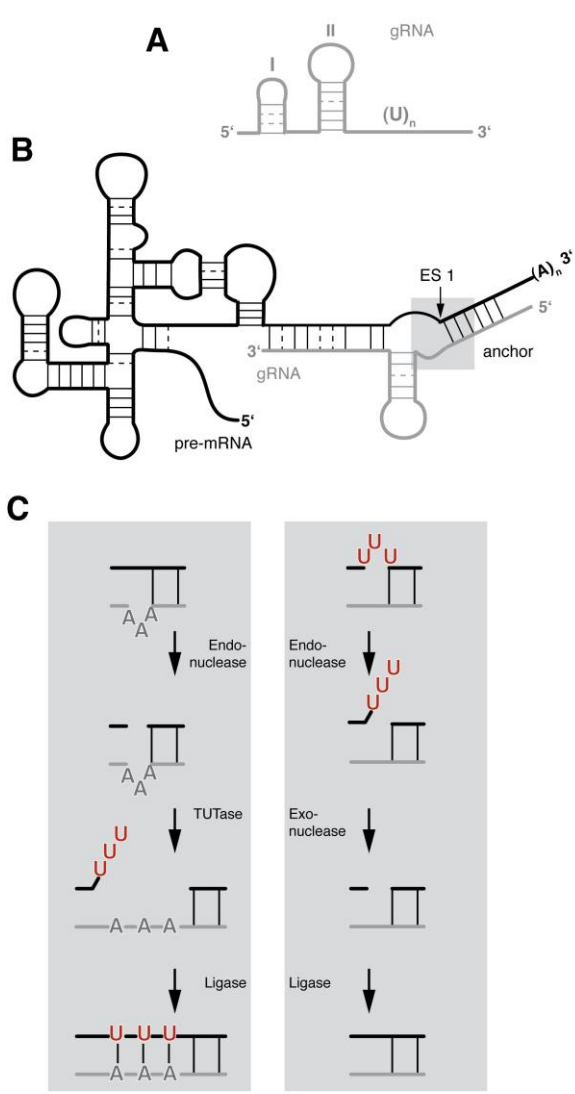
mRNA-gRNA Duplex

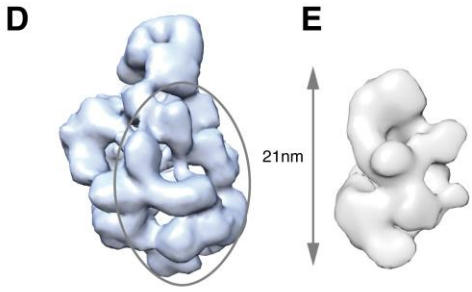
As gRNAs serve as template for RNA uridylyl insertion/deletion RNA editing, faithful annealing to the pre-mRNA is a prerequisite for RNA editing. The gRNA contains a 5'-region that is base complementary to a corresponding region in the pre-edited mRNA next to an editing block, the so-called anchor region. That anchor region lies in the thermodynamically less stable 5' stem I of the free gRNA (Schmid et al., 1995). This stem has to get unwound so that it can anneal to the pre-mRNA. The formed duplex is mainly composed of Watson-Crick base pairs. The second stem (stem II) of the gRNA is retained throughout binding to the mRNA. Stem II contains the so called guiding region. The 3'oligo(U) tail of the gRNA forms base pairs with purine-rich sequences 5' of the editing block. This base pairing further stabilizes the duplex and possibly disrupts mRNA secondary structures, which would otherwise interfere with the binding of the editosome. This common duplex structure was found for different gRNA-pre-mRNA pairs (Figure I.2B) (Leung and Koslowsky, 1999, 2001), but the contribution of the oligo(U) tail to the stability can differ between various pairs (Koslowsky et al., 2004).

The RNA editing is initiated by the endonucleolytic cleavage of the mismatched region of the pre-mRNA next to the anchor duplex. By insertion or deletion of Us, base complementarity of each editing site is accomplished. Throughout the first few RNA editing events the characteristic stem-loop structure is maintained by alternative base pairing of the oligo(U) tail with the mRNA (Yu and Koslowsky, 2006). When RNA editing proceeds through the editing block, the number of duplex base pairs rises and the gRNA stem-loop structure is disrupted in favor of the gRNA /mRNA duplex. While the anchor duplex is characterized by mainly Watson-Crick base pairs, the rest consists, to a large extent, of G-U-wobble base pairs.

This results in a lower thermal stability of the 5'-region of the editing block, possibly allowing thermal breathing. As RNA editing has a general 3' to 5' processivity, thermal breathing allows for anchor annealing with the next gRNA. This general directionality is determined by the complementarity of upstream gRNA anchor regions with the region that got already edited in the partially edited mRNA (Maslov and Simpson, 1992).

In summary, RNA editing in trypanosomes involves two RNA species, comprising structural elements that are partially of high thermodynamic stabilities. These RNAs have to anneal properly at short base complementary anchor regions prior to the RNA editing initiation, which was shown to be influenced by the pre-mRNA secondary structure (Reifur et al., 2010). After RNA editing of the editing block is completed, a fully base-paired thermodynamically stable gRNA /mRNA duplex emerged. This duplex has to dissociate to enable subsequent gRNA annealing.





F

Name	Conserved structural motif(s)	Function
MP81/KREPA1	Zn finger (2x), OB-fold	Interaction
MP63/KREPA2	Zn finger (2x), OB-fold	Interaction
MP42/KREPA3	Zn finger (2x), OB-fold	Interaction, Endo-/ exo-nuclease (in vitro)
MP24/KREPA4	OB-fold	Interaction, gRNA-binding, annealing
MP19/KREPA5	OB-fold	Interaction
MP18/KREPA6	OB-fold	interaction, gRNA-binding
MP46/KREPB4	U1-like, RNaseIII, Pum	Editosome integrity
MP44/KREPB5	U1-like, RNaseIII, Pum	Editosome integrity
MP49/KREPB6	U1-like	Interaction
MP47/KREPB7	U1-like	Interaction
MP41/KREPB8	U1-like	Interaction
MP90/KREN1	U1-like, RNaseIII, dsRBM	U-deletion site specific endonuclease
MP61/KREN2	U1-like, RNaseIII, dsRBM	U-insertion site specific endonuclease
MP67/KREN3	U1-like, RNaseIII, dsRBM	COXII-specific endonuclease
MP100/KREX1	5'-3' Exo, EEP	Nuclease/ nucleotidyl phosphatase
MP99/KREX2	5'-3' Exo, EEP	Nuclease/ nucleotidyl phosphatase
MP57/KRET2	PAP cat/assoc.	TUTase
MP52/KREL1	Ligase, tau, K	RNA ligase
MP48/KREL2	Ligase, tau, K	RNA ligase

Figure I.2. RNA Editing in *T. brucei*.
A. Scheme of gRNA secondary structure. I and II indicate the two stem loops, with I being the less stable one harboring the anchor-region.
B. gRNA annealed to the first anchor region of a structured pre-mRNA. ES 1 indicates the first editing site.
C. Enzymatic steps of the U-insertion and U-deletion RNA editing cycle that is performed at each editing site.
D. Cryo electron microscopic structure of the 35-40S editosome (Golas et al., 2009). The ellipse indicates the region where the 20S editosome can be modeled into the structure.
E. Cryo electron microscopic structure of the 20S editosome (Golas et al., 2009).
F. Stably associated protein inventory of the 20S editosome (modified from Göringer, 2012). Dark gray: proteins (predicted to be) involved in editosome structure, light gray: enzymatic active proteins involved in the editing cycle.
Abbreviations: OB-fold, oligosaccharide/oligonucleotide-binding fold; U1-like, U1-like Zn-finger domain; Pum, Pumilio domain; tau, microtubule-associated tau motif; K, kinesin light-chain domain; PAP, poly(A) polymerase catalytic and associated (cat/assoc.) domains; dsRBM, double-stranded RNA-binding motif; 5′−3′ Exo, 5′−3′ exoribonuclease domain; Zn finger, C2H2-type Zn-finger domain; EEP, endonuclease-exonuclease-phosphatase domain; TUTase, terminal uridylyl transferase.

RNA Structure Remodeling Proteins Involved in U-Insertion/Deletion-Type RNA Editing

The RNA structure remodeling processes can be assisted by proteins. Different classes of structure remodeling proteins were found to be involved in mitochondrial RNA editing in *T. brucei*. These proteins are presented in the following sections.

RNA Helicases Involved in U-Insertion/Deletion-Type RNA Editing

RNA Editing Helicase mHel61p

A search for DEAD-box protein genes in *T. brucei* identified a single-copy gene (Tb927.11.8870), encoding a 61kDa (547 amino acid) polypeptide with a pI of 8.3 (Missel et al., 1997). The encoded polypeptide was named mHel61p and compared to other sequences of the DEAD-box family. The trypanosomal protein shares 29% identity with the initiation factor eIF4A of *Saccharomyces cerevisiae* (Foreman et al., 1991) and 31% identity with a DEAD-box helicase of the hyperthermophilic organism *Methanococcus jannaschii* (Story et al., 2001). Among the kinetoplastida, the identity is even higher. The orthologue proteins share 60% identity between *T. brucei* and *Leishmania major* (LmjF28.2080) and 68% identity between *T. brucei* and *Trypanosoma cruzi* (TcCLB.511801.60). The paralogous genes are localized on non-syntenic chromosomes: chromosome 28 in *L. major* and chromosome 11 in *T. brucei*. The orthologues in *Leishmania* and *T. cruzi* have similar molecular masses with pIs of 8.7 and 7.9. The helicase mHel61p has a putative mitochondrial import sequence, and was experimentally verified as a mitochondrial matrix protein (Missel et al., 1997, Li et al., 2011).

T. brucei mHel61p is expressed as a 2.3-kb transcript in both, the insect and the bloodstream lifecycle stages of the parasite. Single-allele, insect stage knockout trypanosomes behave indistinguishably from wild-type parasites, and *mhel61* null trypanosomes were found to be viable. However, they multiply with a significantly reduced growth rate (Missel et al., 1997), which indicates that *mhel61* is not essential in African trypanosomes, although the absence of the protein causes a metabolic deficiency.

On a molecular level, *mhel61* null mutants are characterized (at steady-state conditions) by strongly reduced amounts of edited mRNAs, while never edited mitochondrial mRNAs and nuclear transcripts are unaffected. This phenotype can be completely rescued by ectopically re-expressing mHel61p in the knockout parasite cell line (Missel et al., 1997). However, mHel61p does not seem to be an integral component of the 20S editosome (Missel et al., 1997). The majority of the polypeptide was identified as a non-complexed mitochondrial protein and only small amounts of mHel61p were found in fractions with an apparent S-value of 20S or higher. Among several possible scenarios, this suggests that mHel61p interacts only transiently, during a specific step of the reaction cycle, with the RNA editing machinery.

Support for this interpretation was gained from the result that the RNA editing *in vitro* activity was not impaired in mitochondrial extracts derived from the mHel61p null mutant. No difference, neither qualitatively nor quantitatively, was observed with respect to the formation of the correct RNA editing product. This implies that editosomes are properly assembled in *mhel61* null trypanosomes and that they function indistinguishably from the processing machinery in wild-type mitochondria. As a consequence, the involvement of mHel61p in the RNA editing process is indirect, functioning in a post-catalytic event that is not monitored in the "single-round" *in vitro* RNA editing experiments.

Further support for these data were published by Li et al. (Li et al., 2011). The authors describe a reduction in the steady-state levels of some edited RNAs in a transgenic *T. brucei mhel61* RNA interference (RNAi) strain. The effect was more pronounced for some mitochondrial transcripts such as CR3 and A6 but rather weak for other pan-edited or moderately edited pre-mRNAs. Whether that reflects a substrate specificity of the protein, or simply an experimental variation, remains to be verified. However, the authors gained indirect evidence for a possible function of mHel61p by performing a qRT-PCR analysis using primer pairs to detect partially edited transcripts of A6 and CYb. By analyzing pre-edited sequence domains that rely on overlapping gRNAs, they identified a progressive decrease in the transcript abundance with increasing distance from the 3'-end of the two mRNAs. In other words: in the absence of mHel61p, upstream RNA editing domains are less frequently edited. This suggests that mHel61p is involved in the succession of RNA editing from one gRNA-mediated editing domain to the next. Two explanations can account for the observed phenomenon: (1) mHel61p is involved in the unwinding of mRNA secondary and/or tertiary structure elements, which enables gRNAs to access their annealing site or (2) the protein mediates the displacement of a downstream gRNA to allow the annealing of the next gRNA. Both scenarios imply that mHel61p must execute an archetypical RNA helicase activity, which was verified by Li et al. (Li et al., 2011) using a recombinant version of *Leishmania major* mHel61p termed LmREH1.

Immunolocalization experiments confirmed the mitochondrial localization of ectopically expressed, tagged mHel61p (Li et al., 2011). In addition to a diffuse distribution over the entire mitochondrial network, some mHel61p was identified in close proximity to the kinetoplast, together with the editosome core polypeptide REL1 and ribosomal proteins. Interestingly, a similar distribution was described for proteins of the gRNA binding complex MRB1, which contains the RNA editing helicase REH2 (Hashimi et al., 2009). These data suggest that RNA editing and possibly translation preferentially occur close to the site of transcription.

RNA Editing Helicase 2 (REH2)

Evidence for a second mitochondrial RNA helicase potentially involved in RNA editing was derived from co-immunoprecipitation and co-purification studies in *Leishmania* and *T. brucei* using several monoclonal antibodies and TAP-tagged proteins believed to be involved in gRNA metabolism (Panigrahi et al., 2008, Weng et al., 2008, Hashimi et al., 2008). The identified polypeptide is 2167 amino acids long (241kDa) and has been named RNA Editing Helicase 2 (REH2; Tb927.4.1500). REH2 has a C-terminal helicase domain that contains a DExH-motif and a long N-terminal extension containing a double strand RNA binding domain (dsRBD). The protein is essential in insect stage trypanosomes: RNAi-mediated knockdown of REH2 causes cells to cease growth after 5 days (Hashimi et al., 2009). An analysis of the steady-state levels of mitochondrial transcripts revealed a reduction of some edited transcripts, without accumulation of their pre-edited forms, while other mRNAs (COII and ND7) were not affected. In addition, REH2-knockdown cells show reduced gRNA levels as well as a disruption of the MRB1 complex (Ammerman et al., 2012, reviewed in Hashimi et al., 2013), suggesting an involvement of REH2 in the biogenesis of gRNAs and in RNA editing.

Co-immunoprecipitation experiments with REH2-specific antibodies verified an ATP-dependent RNA unwinding activity for the REH2 using an A6 mRNA-derived partially double-stranded substrate RNA (72nt/76nt) or a dsRNA molecule (27bp) with 3' overhangs (Hernandez et al., 2010). Fractionation of mitochondrial lysates in glycerol gradients followed by RNA helicase assays and Western blot detection demonstrated that the RNA unwinding activity cosediments with REH2. Furthermore, a photo-crosslinking experiment, using an RNA substrate containing a photo-reactive thio-U residue at the processing site, identified REH2 in close proximity (≤4Å) to the editing site. Although additional experiments are required to corroborate this observation, the result localizes REH2 right at the catalytic center of the RNA editing machinery, which suggests that the protein might contribute to the local unwinding of the pre-RNA/gRNA hybrid at the editing site.

The physical interaction of REH2 with the RNA editing machinery was confirmed using several lines of evidence (Hernandez et al., 2010). Western blot experiments of "native" editosome preparations (*i.e.* purified by conventional column chromatography from mitochondrial detergent lysates) identified REH2 as a binding partner of 20S editosomes although in substoichiometric amounts. Obviously, within the steady-state ensemble of editosomes, not every complex is associated with a REH2 molecule. Furthermore, the isolation of mitochondrial complexes using a TAP-tagged version of REH2 co-purified the majority of the 20S editosome protein inventory. REH2 was also detected in antibody-mediated pulldown experiments using several of the editosome core proteins as a bait (Hernandez et al., 2010). In most of these cases the yield was very low, which was attributed to a low accessibility of the affinity tags. In addition, several editosome core proteins were shown to co-purify with ectopically expressed, TAP-tagged REH2. Interestingly, among the proteins copurifying by immunoprecipitation with anti-REH2 antibodies was mHel61p and another DExH-box containing protein (Tb927.4.3020). Bioinformatic search tools predict a mitochondrial localization for Tb927.4.3020, however, the molecule has not been analyzed to date.

The association of REH2 with protein components of the 20S editosome was further verified in sedimentation experiments of mitochondrial lysates: A subpopulation of REH2 forms complexes >20S, which co-sediment with editosomal proteins. Importantly, the interaction is sensitive to nuclease treatment (RNase A, RNase T1, Micrococcal nuclease), supporting the hypothesis that the interaction between REH2 and the editosome is RNA-mediated utilizing the dsRBD domain of the protein.

RNA Annealers Involved in U-Insertion/Deletion-Type RNA Editing

gBP21

Experiments, utilizing UV cross-linking of proteins to gRNAs, identified a protein with a molecular mass of 21kDa associated with these RNAs (Köller et al., 1994). The protein, called gBP21, is rich in positively charged amino acids. It binds gRNAs with nanomolar dissociation constants, independent of the nucleotide sequence. The binding of gBP21 conserves and stabilizes the structure of large parts of the gRNA (Köller et al., 1997). That interaction especially protects nucleotides at the base of the gRNA stem II against endonucleolytic digestion (Hermann et al., 1997). Immunoprecipitation of gBP21 showed partial association with the RNA editing complex (Allen et al., 1998), but the largest amount of gBP21 is present in a free form. As knockout studies showed, the protein is not essential for RNA editing, but is likely involved in multiple mitochondrial processes. Only cells of the bloodstream form, which have a reduced mitochondrial function, stay viable after the knockout (Lambert at al., 1999).

The protein gBP21 was shown to stimulate the annealing of partially complementary mRNA-gRNA duplexes *in vitro*. This annealing activity of the protein results in a thirtyfold stimulation of the second order rate constant of the RNA annealing and is independent of ATP. It is rather unspecific as it also stimulates the annealing of unrelated RNA duplexes and RNA-DNA hybrids, but the efficiencies are dependent on the presence of the gRNA stem-loop II structure (Müller et al., 2001). The mechanism of gBP21 activity was investigated by Müller and Göringer in 2002 and a detailed mechanism was proposed, whereby the protein acts as matchmaker by forming a kinetically stable 1:1 complex with the gRNA. Thereby it unwinds the 5'-hairpin and "presents" the anchor sequence. The binding of gRNA involves up to six ionic bonds contributing about 75%

of the total free energy of the interaction. This ionic interaction neutralizes the charge of the gRNA backbone, and thereby the charge repulsion to the complementary mRNA, and enhances the collision frequency of the two RNAs. After association of the gRNA with its partially complementary pre-mRNA, the protein dissociates due to its two orders of magnitude lower affinity for double-stranded RNA (Müller and Göringer, 2002). This mechanism was later verified by a crystal structure of the gBP21/ gBP25/gRNA complex (see below) (Schumacher at al., 2006).

TbRGG2

TbRGG2 is a mitochondrial protein, which has an N-terminal glycine-rich domain comprising an RGG-motif and a C-terminal RNA recognition motif. It co-immunoprecipitates with editosomes, but is not associated with editosomes purified by multiple-step biochemical purification (Panigrahi et al., 2003). It can bind RNA with a preference for poly(U) and is essential in both the mammalian and insect lifecycle stages, with a knockdown leading to a specific decrease of pan-edited RNAs (Fisk et al., 2008). Detailed sequence analysis of the RNAs in knockdown strains showed that the progression of RNA editing from 3' to 5' is impacted on pan-edited transcripts, indicating that it is involved in the efficiency of subsequent annealing of gRNAs. It has an *in vitro* RNA annealing activity and can act *in vivo* as transcription anti-terminator in *E. coli* cells (Ammerman et al., 2010). The protein has a low nM affinity for the pre-edited RPS12 transcript while the dissociation constant for edited transcripts is by an order of magnitude higher and still slightly higher for gRNA. Construction of shortened protein variants showed that the N-terminal G-rich domain is the mediator of high affinity RNA-binding and *in vitro* RNA annealing activity, while the C-terminal RNA recognition motif performs the RNA unwinding activity. However, complementation studies showed that the RNA unwinding activity is not essential for cell viability (Foda et al., 2012).

The Mitochondrial RNA Binding Complex

Another mitochondrial protein gBP25 gained attention due to its homology to proteins in other kinetoplastids (Blom et al., 2001). Its orthologue in *Leishmania tarentolae* is in a complex with the gBP21 orthologue. That complex performs RNA annealing (Aphasizhev et al., 2003). The interdependence of gBP21 and gBP25 protein stability in knockdown experiments also indicated an association of the two proteins in *T. brucei* (Vondrusková et al., 2005). Schumacher at al. (2006) crystallized a gBP25/gBP21 complex to yield an atomic resolution structure. Although the two proteins have nearly no sequence identity, both form a nearly identical "whirly" fold that is mainly composed of ß-sheets, with four monomers arranged in a hetero-tetramer. The hetero-tetramer binds gRNAs with nanomolar dissociation constants, with the gRNA stem loop II being sufficient for high affinity binding. The crystal structure of the gBP21-gBP25 complex, in conjunction with a gRNA, supports the mode of action that was proposed for gBP21 alone (Müller and Göringer, 2002). The binding occurs at a highly electropositive, curved β-sheet surface. Stem II is bound in a binding pocket in a structure conservative fashion, whereas gRNA stem I is unwound to present the anchor region. The structure of the protein complex is unaltered throughout RNA binding (Schumacher et al., 2006). The complex does not co-purify with editosomes after affinity purification (Zíková et al. 2008).

In the last years an image emerged, where multiple proteins associate to form a mitochondrial RNA binding (MRB1) complex (reviewed in Hashimi et al., 2013, Aphasizheva and Aphasizhev, 2016). A core of this complex, composed of six different proteins, includes the described gBP21-gBP25 complex. Multiple other proteins are less stably associated to this core, with a subcomplex containing TbRGG2, and REH2 being associated to the MRB. Additionally to the transient association of MRB components with the editosome, interactions with proteins involved in other processing steps were found, including proteins involved in polyadenylation/uridylation and the large subunit of the ribosome (Aphasizheva et al., 2011). On basis of that, two functions of the MRB1 complex were proposed: mediation of the recruitment and exchange of gRNAs that are required for multiple rounds of mitochondrial RNA editing, especially in pan-edited transcripts and linkage of RNA editing to other RNA processing steps (Hashimi et al., 2013).

RBP16

RBP16 is a 16 kilodalton protein with a predicted N-terminal cold shock domain and an RGG-motif. It was first purified from mitochondrial extracts due to its affinity for poly(U). Gel shift assays of RBP16 with gRNA revealed that micromolar concentrations are needed for a stable association. Despite this low affinity, about 30% of gRNAs and also 30% of 9S and 12S rRNAs were found to co-immunoprecipitate with the protein (Hayman and Read, 1999). The salt independence of its RNA binding suggests a subsidiary role of electrostatic interactions. Binding studies revealed a minimal RNA binding substrate consisting of a stretch of four Us with 14nt and 4nt extensions at the 5'- and 3'-ends, respectively. Importantly, RBP16 also binds to a pre-mRNA with a similar affinity as for gRNA (Pelletier et al., 2000). This indicates that it is a relatively unspecific RNA binding protein. Neither affinity purification nor immunoprecipitation revealed an association of RBP16 with the editosome or any gRNA binding protein (Hayman et al., 2001). Knockdown studies showed that RBP16 is essential in procyclic *T. brucei* and alters the stability of mitochondrial transcripts. The efficiency of RNA editing is only altered in the differentially edited CYb transcripts (Pelletier and Read, 2003). The result indicates that the protein is not particularly involved in RNA editing but has a

multifunctional role in the mitochondrion. Despite its low effects on RNA editing *in vivo*, RBP16 was shown to enhance the first step in the insertion editing reaction cycle *in vitro* (Miller et al., 2006) congruent with an *in vitro* RNA annealing activity. It is annealing-active both on cognate mRNA/gRNA pairs and on unrelated complementary RNAs, with no need for an oligo(U) tail. RBP16 was found to also have an RNA chaperone-like RNA unwinding activity, as it performs transcription anti-termination in *E. coli* cells (Ammerman et al., 2008).

Most of the described mitochondrial proteins, that were found to have an RNA annealing activity, are largely promiscuous for different substrates *in vitro*. *In vivo* TbRGG2 has a specific effect on efficient RNA editing progression throughout the mRNA, while the other proteins seem to have more multifunctional roles in the mitochondrion. There likely is redundancy with different proteins doing similar jobs, as simultaneous knockdown of the gBP21-gBP25 complex and RBP16 leads to enhanced defects in RNA editing for some edited RNAs (Fisk et al., 2009).

Chaperones Involved in U-Insertion/Deletion-Type RNA Editing

Böhm et al. 2012 found that the 20S editosome has an intrinsic ATP-independent RNA unwinding activity. This activity leads to an increased accessibility of different never- and pre-edited transcript for both a single-strand-RNA-specific and a double-strand-RNA-specific endonuclease, suggesting a general "loosening" affecting the tertiary structure of the RNA. Atomic force microscopic images showed that the RNA becomes unwound to an elongated form by the RNA chaperone activity of the 20S editosome (Böhm et al., 2012). This activity was investigated, utilizing chemical probing, to map the structural changes, which 20S editosomes induce at the single nucleotide positions in pre-mRNAs. The experiments show that 20S editosomes act by generally destabilizing the secondary structure of the RNA, which results in a simplified RNA folding landscape with reduced energy barriers for gRNA binding. The complex acts qualitatively and quantitatively similar on never edited and pre-edited RNAs. The RNA chaperone activity results in enhanced flexibilities of preferentially U-nucleotides (Leeder et al. 2016a). Until now it is not known which of the integral proteins of the 20S editosome is responsible for the RNA chaperone activity. Six proteins containing an OB-fold were discussed as possible candidates (Böhm et al., 2012, Leeder et al., 2016a) (see chapter III).

Research Aim

As a consequence of the structure-function paradigm of biomolecules ("structure determines function"), any understanding of a biological process must involve the molecular description of the dynamic alterations of the participating biomolecules. Within this context, RNA structure remodeling reactions have recently attracted special attention, because RNA molecules are involved in nearly every cellular process. This includes the mitochondria-specific RNA editing reaction in African trypanosomes and other protozoan parasites. RNA editing requires a remodeling of the structurally highly stable pre-edited substrate mRNAs to allow for the binding of cognate gRNA molecules as a first step in the reaction cycle.

The experimental work presented here aims at unraveling the molecular nature of the editosome-intrinsic RNA remodeling activity. For that, I developed an assay system, which is capable to detect and characterize structural changes in RNA molecules (**chapter II)**. To identify the editosomal proteins that catalyze the RNA unwinding reaction, I focused on the six OB-fold proteins of the editosome. This is presented in **chapter III**. I expressed the different proteins as recombinant polypeptides and after purification I analyzed their RNA chaperone activity. I present first experimental evidence that the OB-fold proteins of the *T. brucei* editosome execute RNA unfolding activity. Furthermore, I provide a first molecular model of the 20S editosome, which integrates the generated biochemical data into a coherent structural context. In **chapter IV**, I further characterize the RNA remodeling activity of an individual OB-fold protein especially in comparison to the activity of the fully assembled editosome.

Differential RNA editing is observed throughout the lifecycle of *T. brucei*. However, no major changes in the composition of the editosome or in the abundance of gRNAs have been detected. This has led to the hypothesis, that changes in the mitochondrial environment might influence the architecture and/or activity of the editosome. Since the editosomal OB-fold proteins are, aside from their functional importance, also of structural importance, the aim of the experiments, described in **chapter V**, is to explore differential protein complex assembly in changed environments.

References

Alemán, E.A., Lamichhane, R., and Rueda, D. (2008). Exploring RNA folding one molecule at a time. Curr. Opin. Chem. Biol. 12, 647-654.

Allen, T.E., Heidmann, S., Reed, R., Myler, P.J., Göringer, H.U., and Stuart, K.D. (1998). Association of Guide RNA Binding Protein gBP21 with Active RNA Editing Complexes in *Trypanosoma brucei*. Mol. Cell. Biol. 18, 6014-6022.

Ammerman, M.L., Fisk, J.C., and Read, L.K. (2008). gRNA/pre-mRNA annealing and RNA chaperone activities of RBP16. RNA 14, 1069-1080.

Ammerman, M.L., Presnyak, V., Fisk, J.C., Foda, B.M., and Read, L.K. (2010). TbRGG2 facilitates kinetoplastid RNA editing initiation and progression past intrinsic pause sites. RNA 16, 2239-2251.

Ammerman, M.L., Downey, K.M., Hashimi, H., Fisk, J.C., Tomasello, D.L., Faktorová, D., Kafková, L., King, T., Lukes, J., and Read, L.K. (2012). Architecture of the trypanosome RNA editing accessory complex, MRB1. Nucleic Acids Res. 40, 5637-5650.

Aphasizhev, R., Aphasizheva, I., Nelson, R.E., and Simpson, L. (2003). A 100-kD complex of two RNA-binding proteins from mitochondria of *Leishmania tarentolae* catalyzes RNA annealing and interacts with several RNA editing components. RNA 9, 62-76.

Aphasizheva, I., Maslov, D., Wang, X., Huang, L., Aphasizhev, R. (2011). Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. Mol Cell. 42, 106-117.

Aphasizhev, R., and Aphasizheva, I. (2011). Mitochondrial RNA processing in trypanosomes. Res. Microbiol. 162, 655-663.

Aphasizheva, I., and Aphasizhev, R. (2016). U-Insertion/Deletion mRNA-Editing Holoenzyme: Definition in Sight. Trends Parasitol. 32, 144-156.

Blom, D., Burg, J., Breek C.K., Speijer, D., Muijsers, A.O., and Benne, R. (2001). Cloning and characterization of two guide RNA-binding proteins from mitochondria of *Crithidia fasciculata*: gBP27, a novel protein, and gBP29, the orthologue of *Trypanosoma brucei* gBP21. Nucleic Acids Res. 29, 2950-2962.

Blum, B., and Simpson, L. (1990). Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. Cell 62, 391-397.

Böhm, C., Katari, V.S., Brecht, M., and Göringer, H.U. (2012). *Trypanosoma brucei* 20 S Editosomes Have One RNA Substrate-binding Site and Execute RNA Unwinding Activity. J. Biol. Chem. 287, 26268-26277.

Cech, T.R., and Steitz, J.A.(2014). The Noncoding RNA Revolution - Trashing Old Rules to Forge New Ones. Cell 157, 77-94.

Crick, F.H. (1958). On Protein Synthesis. Sym. Soc. Exp. Biol. 12, 138-163.

Crick, F.H.C. (1968). The origin of the genetic code. J. Mol. Biol. 38, 367-379.

Czerwoniec, A., Kasprzak, J.M., Bytner, P., Dobrychłop, M., and Bujnicki, J.M. (2015). Structure and intrinsic disorder of the proteins of the *Trypanosoma brucei* editosome. FEBS Lett. 589, 2603-2610.

Ditzler, M.A., Rueda, D., Mo, J., Håkansson, K., and Walter, N.G.(2008). A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. Nucleic Acids Res. 36, 7088-7099.

Draper, D. (1992). The RNA-Folding Problem. Acc. Chem. Res. 25, 201-207.

Endoh, T., Kawasaki, Y., and Sugimoto, N. (2013). Suppression of Gene Expression by G-Quadruplexes in Open Reading Frames Depends on G-Quadruplex Stability. Angew. Chem. Int. Ed. Engl. 52, 5522-5526.

Etheridge, R.D., Aphasizheva, I., Gershon, P.D. and Aphasizhev, R. (2008). 3' adenylation determines mRNA abundance and monitors completion of RNA editing in *T. brucei* mitochondria. EMBO J. 27, 1596-1608.

Fisk, J.C., Ammerman, M.L., Presnyak, V., and Read, L.K. (2008). TbRGG2, an essential RNA editing accessory factor in two *Trypanosoma brucei* life cycle stages. J. Biol. Chem. 283, 23016-23025.

Fisk, J.C., Presnyak., V., Ammerman, M.L., and Read, L.K. (2009). Distinct and Overlapping Functions of MRP1/2 and RBP16 in Mitochondrial RNA Metabolism. Mol. Cell. Biol. 29, 5214-5225.

Foda, B.M., Downey, K.M., Fisk, J.C., and Read, L.K. (2012). Multifunctional G-Rich and RRM-Containing Domains of TbRGG2 Perform Separate yet Essential Functions in Trypanosome RNA Editing. Eukaryotic Cell 11, 1119-1131.

Foreman, P.K., Davis, R.W., and Sachs, A.B. (1991). The *Saccharomyces cerevisiae* RPB4 gene is tightly linked to the TIF2 gene. Nucleic Acids Res. 19, 2781.

Gerczei, T., and Correll, C. (2004). Imp3p and Imp4p mediate formation of essential U3-precursor rRNA (pre-rRNA) duplexes, possibly to recruit the small

subunit processome to the pre-rRNA. Proc. Natl. Acad. Sci. USA 101, 15301-15306.

Gilbert, W. (1986). Origin of life: The RNA world. Nature 319, 618.

Göringer, H.U. (2012). 'Gestalt,' composition and function of the *Trypanosoma brucei* editosome. Annu. Rev. Microbiol. 66, 65-82.

Golas, M.M., Böhm, C., Sander, B., Effenberger, K., Brecht, M., Stark, H., and Göringer, H.U. (2009). Snapshots of the RNA editing machine in trypanosomes captured at different assembly stages *in vivo*. EMBO J. 28, 766-778.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell. 35, 849-857.

Hashimi, H., Zíková, A., Panigrahi, A.K., Stuart, K.D., and Lukeš, J. (2008). TbRGG1, an essential protein involved in kinetoplastid RNA metabolism that is associated with a novel multiprotein complex. RNA 14, 970-980.

Hashimi, H., Cicová, Z., Novotná, L., Wen, Y.Z., and Lukeš, J. (2009). Kinetoplastid guide RNA biogenesis is dependent on subunits of the mitochondrial RNA binding complex 1 and mitochondrial RNA polymerase. RNA 15, 588-599.

Hashimi, H., Zimmer, S.L., Ammerman, M.L., Read, L.K., and Lukeš, J. (2013). Dual core processing: MRB1 is an emerging kinetoplast RNA editing complex. Trends Parasitol. 29, 91-99.

Hayman, M.L,, and Read, L.K. (1999). *Trypanosoma brucei* RBP16 is a mitochondrial Y-box family protein with guide RNA binding activity. J. Biol. Chem. 274, 12067-12074.

Hayman, M.L., Miller, M.M., Chandler, D.M., Goulah, C.C., Read, L.K. (2001). The trypanosome homolog of human p32 interacts with RBP16 and stimulates its gRNA binding activity. Nucleic Acids Res. 29, 5216-5225.

Hernandez, A., Madina, B.R., Ro, K., Wohlschlegel, J.A., Willard, B., Kinter, M.T., and Cruz-Reyes, J. (2010). REH2 RNA helicase in kinetoplastid mitochondria: ribonucleoprotein complexes and essential motifs for unwinding and guide RNA (gRNA) binding. J. Biol. Chem. 285, 1220-1228.

Hermann, T., Schmid, B., Heumann, H., and Göringer, H.U. (1997). A three-dimensional working model for a guide RNA from *Trypanosoma brucei*. Nucleic Acids Res. 25, 2311-2318.

Herschlag, D., Khosla, M., Tsuchihashi, Z., and Karpel, R.L. (1994). An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. EMBO J. 13, 2913-2924.

Herschlag, D. (1995). RNA chaperones and the RNA folding problem. J. Biol. Chem. 270, 20871-20874.

Jensen, R.E., and Englund, P.T. (2012). Network News: The Replication of Kinetoplast DNA. Annu. Rev. Microbiol. 66, 473-491.

Kao, C.Y., and Read, L.K. (2005). Opposing effects of polyadenylation on the stability of edited and unedited mitochondrial RNAs in *Trypanosoma brucei* Mol. Cell. Biol. 25, 1634-1644.

Köller, J., Norskau, G., Paul, A.S., Stuart, K., and Göringer, H.U. (1994). Different *Trypanosoma brucei* guide RNA molecules associate with an identical complement of mitochondrial proteins *in vitro*. Nucleic Acids Res. 22, 1988-1995.

Köller, J., Müller, U.F., Schmid, B., Missel, A., Kruft, V., Stuart, K., and Göringer, H.U. (1997). *Trypanosoma brucei* gBP21 an arginine-rich mitochondrial protein that binds to guide RNA with high affinity. J. Biol. Chem. 272, 3749-3757.

Koslowsky, D.J., Reifur, L., Yu, L. and Chen, W. (2004). Evidence for U-tail stabilization of gRNA/mRNA interactions in kinetoplastid RNA Editing. RNA Biol. 1, 28-34.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell 31, 147-157.

Lambert, L., Müller, U.F., Souza, A.E., and Göringer, H.U. (1999). The involvement of gRNA-binding protein gBP21 in RNA editing-an *in vitro* and *in vivo* analysis. Nucleic Acids Res. 27, 1429-36.

Leeder, W.M., Voigt, C., Brecht, M., and Göringer, H.U. (2016a). The RNA chaperone activity of the *Trypanosoma brucei* editosome raises the dynamic of bound pre-mRNAs. Sci. Rep. 6, 19309.

Leeder, W.M., Hummel, N.F., and Göringer, H.U. (2016b). Multiple G-quartet structures in pre-edited mRNAs suggest evolutionary driving force for RNA editing in trypanosomes. Sci. Rep. 6, 29810.

Leeder, W.M., Reuss, A.J., Brecht, M., Kratz, K., Wachtveitl, J., and Göringer HU. (2015). Charge reduction and thermodynamic stabilization of substrate RNAs inhibit RNA editing. PLoS One. 10, e0118940.

Leitão, A.L., Costa, M.C., and Enguita, F.J. (2015). Unzippers, Resolvers and Sensors: A Structural and Functional Biochemistry Tale of RNA Helicases. Int. J. Mol. Sci. 16, 2269-2293.

Leung, S.S., and Koslowsky, D.J. (1999). Mapping contacts between gRNA and mRNA in trypanosome RNA editing. Nucl. Acid. Res. 27, 778-787.

Leung, S.S., and Koslowsky, D.J. (2001). Interactions of mRNAs and gRNAs involved in trypanosome mitochondrial RNA editing: structure probing of an mRNA bound to its cognate gRNA. RNA 7, 1803-1816.

Li, F., Herrera, J., Zhou, S., Maslov, D.A., and Simpson, L. (2011). Trypanosome REH1 is an RNA helicase involved with the 3'-5' polarity of multiple gRNA-guided uridine insertion/deletion RNA editing. Proc. Natl. Acad. Sci. USA 108, 3542-3547.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., Mesirov, J.P., Cech, T.R., and Chang, H.Y. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell 165, 1267-1279.

Marek, M.S., Johnson-Buck, A., and Walter, N.G. (2011). The shape-shifting quasispecies of RNA: One sequence, many functional folds. Phys. Chem. Chem. Phys. 13, 11524-11537.

Maslov, D.A., and Simpson, L. (1992). The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing. Cell 70, 459-467.

Mayer, O., Rajkowitsch, L., Lorenz, C., Konrat, R., and Schroeder, R. (2007). RNA chaperone activity and RNA-binding properties of the *E. coli* protein StpA. Nucleic Acids Res. 35, 1257-1269.

McDermott, S.M., Luo, J., Carnes, J., Ranish, J.A., and Stuart, K. (2016). The architecture of *Trypanosoma brucei* editosomes. Proc. Natl. Acad. Sci. USA. 113, E6476-E6485.

Miller MM, Halbig K, Cruz-Reyes J, Read LK. (2006). RBP16 stimulates trypanosome RNA editing *in vitro* at an early step in the editing reaction. RNA 12, 1292-1303.

Missel, A., Souza, A.E., Nörskau, G., and Göringer, H.U. (1997). Gene disruption of a mitochondrial DEAD box protein in *Trypanosoma brucei* affects edited mRNAs. Mol. Cell. Biol. 17, 4895-4903.

Müller, U.F., Lambert, L., and Göringer, H.U. (2001). Annealing of RNA editing substrates facilitated by guide RNA-binding protein gBP21. EMBO J. 20, 1394-1404.

Müller, U.F., and Göringer, H.U. (2002). Mechanism of the gBP21-mediated RNA/RNA annealing reaction: matchmaking and charge reduction. Nucleic Acids Res. 30, 447-455.

Orgel, L.E. (1968). Evolution of the genetic apparatus. J. Mol. Biol. 38, 381-393.

Panigrahi, A.K., Allen, T.E., Stuart, K., Haynes, P.A., and Gygi, S.P. (2003). Mass spectrometric analysis of the editosome and other multiprotein complexes in *Trypanosoma brucei*. J. Am. Soc. Mass. Spectrom. 14, 728-735.

Panigrahi, A.K., Zíková, A., Dalley, R.A., Acestor, N., Ogata, Y., Anupama, A., Myler, P.J., and Stuart, K.D. (2008). Mitochondrial complexes in *Trypanosoma brucei*: a novel complex and a unique oxidoreductase complex. Mol. Cell. Proteomics 7, 534-545.

Panja, S., and Woodson, S.A. (2012). Hfq proximity and orientation controls RNA annealing. Nucleic Acids Res. 40, 8690-8697.

Pelletier, M., Miller, M.M., and Read, L.K. (2000). RNA-binding properties of the mitochondrial Y-box protein RBP16. Nucleic Acids Res. 28, 1266-1275.

Pelletier, M., and Read, L.K. (2003). RBP16 is a multifunctional gene regulatory protein involved in editing and stabilization of specific mitochondrial mRNAs in *Trypanosoma brucei*. RNA. 9, 457-468.

Pollard, V.W., Rohrer, S.P., Michelotti, E.F., Hancock, K., and Hajduk, S.L. (1990). Organization of Minicircle Genes for Guide RNAs in *Trypanosoma brucei*. Cell 63, 763-790.

Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M.F., Konrat, R., Bläsi, U., and Schroeder, R. (2007). RNA Chaperones, RNA Annealers and RNA Helicases. RNA Biol. 4, 118-130.

Reifur, L., Yu, L.E., Cruz-Reyes, J., Vanhartesvelt, M., and Koslowsky, D.J. (2010).The Impact of mRNA Structure on Guide RNA Targeting in Kinetoplastid RNA Editing. PLoS One 5, e12235.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S, (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. Nature 505, 701-705.

Schmid, B., Riley, G.R., Stuart, K., and Göringer, H.U. (1995). The secondary structure of guide RNA molecules from *Trypanosoma brucei*. Nucleic Acids Res. 23, 3093-3102.

Schnaufer, A., Wu, M., Park, Y.J., Nakai, T., Deng, J., Proff, R., Hol, W.G., and Stuart, K.D. (2010). A Protein-protein interaction map of trypanosome ~20S editosomes. J. Biol. Chem. 285, 5282-5295.

Schroeder, R., Barta, A., and Semrad, K. (2004). Strategies for RNA folding and assembly. Nature 5, 908-919.

Schultes, E. A., and Bartel, D. P. (2000). One sequence, two ribozymes: implications for the emergence of new ribozyme folds. Science 289, 448-452.

Schumacher, M.A., Karamooz, E., Zíková, A., Trantírek, L., and Lukes, J. (2006). Crystal Structures of *T. brucei* MRP1/MRP2 Guide-RNA Binding Complex Reveal RNA Matchmaking Mechanism. Cell 126, 701-711.

Semrad, K. (2011). Proteins with RNA Chaperone Activity: A World of Diverse Proteins with a Common Task - Impediment of RNA Misfolding. Biochem. Res. Int. 2011, 532908.

Story, R.M., Li, H., and Abelson, J.N. (2001). Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*. Proc. Natl. Acad. Sci. USA 13, 1465-1470.

Suematsu, T., Zhang, L., Aphasizheva, I., Monti, S., Huang, L., Wang, Q., Costello, C.E., and Aphasizhev, R. (2016). Antisense Transcripts Delimit Exonucleolytic Activity of the Mitochondrial 3' Processome to Generate Guide RNAs. Mol. Cell 61, 364-378.

Tompa, P., and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB J. 18, 1169-1175.

Vondrusková, E., van den Burg, J., Zíková, A., Ernst, N.L., Stuart, K., Benne, R., and Lukes, J. (2005). RNA interference analyses suggest a transcript-specific regulatory role for mitochondrial RNA-binding proteins MRP1 and MRP2 in RNA editing and other RNA processing in *Trypanosoma brucei*. J. Biol. Chem. 280, 2429-38.

Weng, J., Aphasizheva, I., Etheridge, R.D., Huang, L., Wang, X., Falick, A.M., and Aphasizhev, R. (2008). Guide RNA-binding complex from mitochondria of trypanosomatids. Mol. Cell 32, 198-209.

Woodson, S.A. (2010). Taming free energy landscapes with RNA chaperones. RNA Biol. 7, 677-686.

Yu, L.E, and Koslowsky, D.J. (2006). Interactions of mRNAs and gRNAs involved in trypanosome mitochondrial RNA editing: Structure probing of a gRNA bound to its cognate mRNA. RNA 12, 1050-1060.

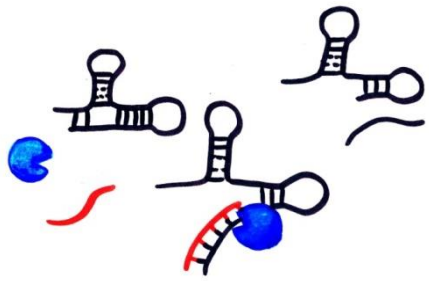
Zíková, A., Kopecná, J., Schumacher, M.A., Stuart, K., Trantírek, L., and Lukes, J. (2008). Structure and function of the native and recombinant mitochondrial MRP1/MRP2 complex from *Trypanosoma brucei*. Int. J. Parasitol. 38, 901-12.

Chapter II

The Editosome-Bound Folding State Favors the Formation of pre-mRNA-gRNA Hybrid RNAs

Results, figures and experimental procedures of this chapter have been included into:

Leeder, W.M., Voigt, C., Brecht, M., and Göringer, H.U. (2016). The RNA Chaperone Activity of the *Trypanosoma brucei* Editosome Raises the Dynamic of Bound pre-mRNAs. Sci. Rep. 6, 19309.



Summary

Trypanosoma brucei editosomes were shown to perform an RNA chaperone-like RNA unwinding of mitochondrial transcripts. This chaperone activity is independent of ATP and results in a structure "loosening" of the RNA by enhancing the flexibility of primarily U-nucleotides. Here I show the development of a biochemical *in vitro* assay that mimics the situation prior to RNA editing. It monitors the effects of the RNA chaperone activity on the entry of a "guiding" DNA-oligonucleotide and subsequent processing steps. The investigation of the RNA chaperone activity of the 20S editosome showed that the activity facilitates an enhanced hybrid formation, which is necessary for subsequent processing. This assay can be used to investigate the dynamic RNA remodeling in detail and it facilitates screening of editosomal proteins for an RNA chaperone activity.

Introduction

Proteins conducting an RNA chaperone activity have been found in many different contexts. Similarly, different assays to probe *in vitro* for an RNA chaperone activity evolved, reviewed by Rajkowitsch et al. (2005). To differentiate between the two general modes of RNA structure remodeling, RNA annealing and strand displacement, the chaperone activity of a protein is investigated in strand displacement and annealing assays on short RNA substrates. These assays were implemented with fluorescent dyes, enabling the use in a time resolved manner. Thus, the acceleration of a single dynamic reaction, by a protein with RNA chaperone activity, can be monitored.
There are also more complex and specialized assay systems, which monitor refolding of larger RNA substrates involving multiple steps. These assays involve self-cleavage of the faithfully folded RNA, by the use of ribozymes or the group I intron as substrate. This subsequent cleavage captures the correctly folded product by making the reaction virtually irreversible.
We investigated the effects of 20S editosome binding on its substrate pre-mRNA and found that specifically Us gain in flexibility (Leeder et al., 2016). Within the scope of that study, I developed a biochemical assay that unravels which implications this increase in flexibility has for the initial steps in RNA editing (Leeder et al. 2016). It shows that the RNA chaperone activity of the 20S editosome alters the mRNA folding landscape to enable gRNA-pre-mRNA hybrid formation. Furthermore, I provide a method to investigate RNA chaperone activity of single proteins involved in RNA editing in *T. brucei*.

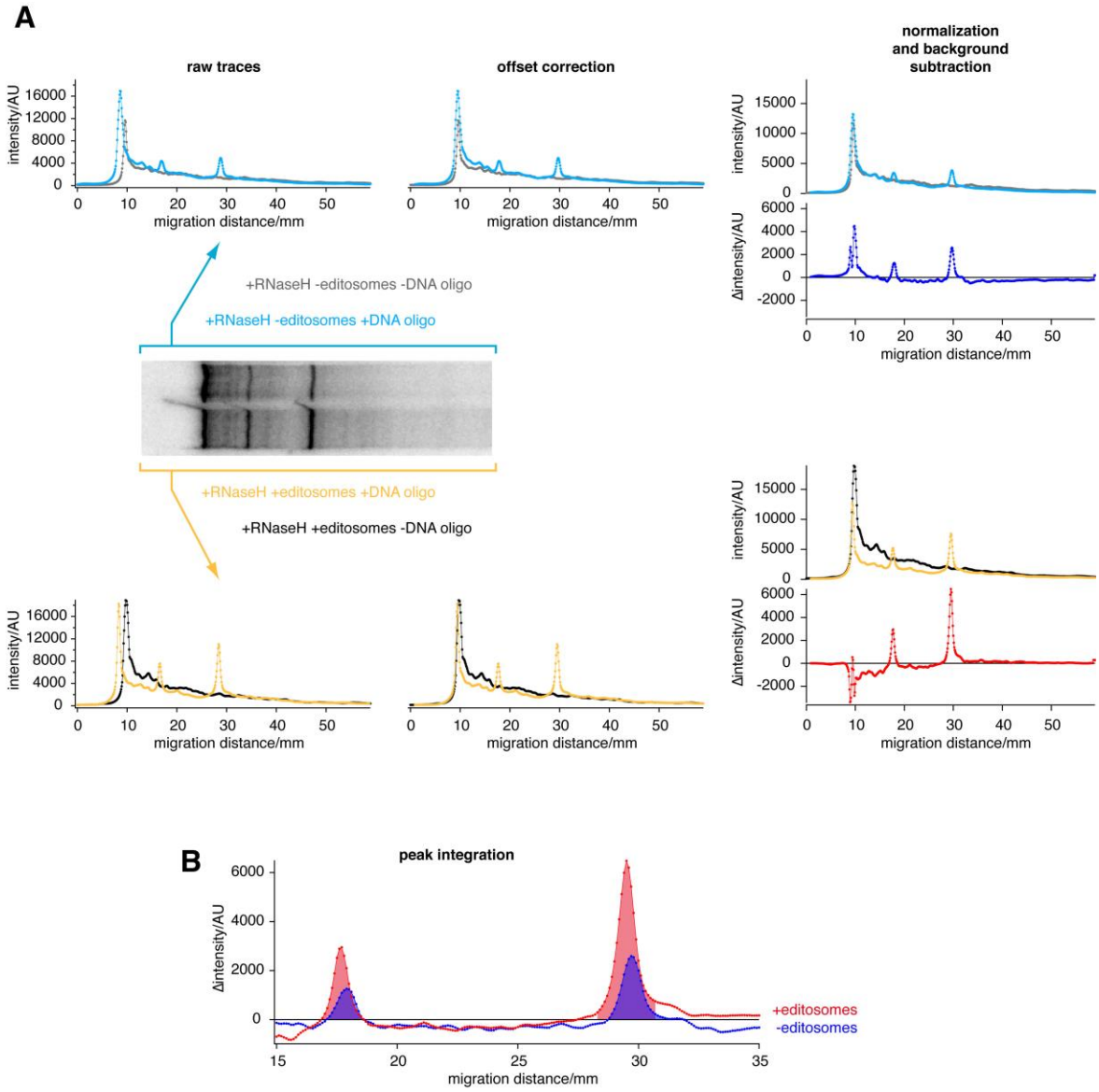


Figure II.1. Example of the Analysis of an RNaseH-Based gDNA Annealing Assay +/- 20S Editosomes.
A. Initial phosphorimaging output, raw densitometry traces, offset correction and background subtraction/normalization.
B. Peak integration: The area of the peaks is used as measure for the amount of the corresponding cleavage product. Afterwards the peak area is normalized to the sample without editosomes.

Results

The mechanistic rationale for increasing the flexibility of editosome-bound RNAs is to overcome the structural rigidity of the different pre-edited transcripts to lower the energy barrier for subsequent steps in the RNA editing cycle. Based on our current knowledge, the next step in the reaction is the annealing of a gRNA molecule to generate a pre-mRNA-gRNA hybrid, which defines the pre-mRNA endonucleolytic cleavage site for the U-insertion/deletion reaction. As a consequence of the data showing enhanced pre-mRNA flexibility after editosome binding, the formation of gRNA-pre-mRNA hybrid RNAs should be thermodynamically favored if the pre-mRNA is in its "flexible" *i.e.* editosome-bound state and disfavored if the transcript is in its "rigid" (free) configuration. To experimentally falsify this hypothesis I developed an assay in which the formation of the pre-mRNA-gRNA hybrid is mimicked by short DNA oligonucleotides as guiding molecules. A mechanistic scheme of the assay is shown in chapter III (Figure III.4) and an example of the data analysis is shown in Figure II.1. This assay enabled me to quantitatively assess hybrid formation by RNaseH cleavage. Representative results are shown in Figure II.2 using CYb as a target pre-mRNA. Invariably, hybrid formation is favored if the transcript is in the editosome-bound folding state and not in the free RNA conformation. Depending on the guiding DNA oligonucleotide, up to 5-fold differences were measured between the two folding states, which demonstrates that at least one function of the editosome chaperone activity is to alter the pre-mRNA folding landscape to promote the formation of gRNA-pre-mRNA hybrid RNAs.

Discussion

The invention of efficient *in vitro* assays accelerated the elucidation of the enzymatic functions of the *T. brucei* editosome, with the pre-cleaved RNA editing assay leading the way (Igo et al., 2000). So the enzymatic components that perform the reactions – from endonucleolytic cleavage, over uridylyl deletion and insertion, to ligation – were discovered and investigated in detail. Although even much earlier the involvement of an RNA "chaperone" function, mediated by an RNA helicase, was discussed (Missel and Göringer, 1994), till this day the dynamic RNA editing reaction steps are largely unexplored.

The assay presented here mimics the situation in the first steps prior to the RNA editing reaction cycle, where the pre-mRNA structure has to be "loosened" for efficient gRNA annealing to occur. The faithfully annealed mRNA-gRNA duplex is the substrate for the subsequent RNA editing reaction, starting with endonucleolytic cleavage, which makes the steps irreversible and opens a way out of the dynamic process.

The results show that the 20S editosome performs RNA chaperone activity that raises the probability for gRNA annealing to its target site. This is a prerequisite for the initiation of RNA editing. This assay provides the opportunity to investigate this activity on different substrate RNAs, at different positions in the RNA and with different proteins. As it is not known to date which protein component executes the RNA chaperone activity, now an assay is available to test the single proteins of the editosome.

Experimental Procedures

Cloning and RNA synthesis

The mitochondrial gene encoding apocytochrome b (CYb) was PCR-amplified from *T. brucei* Lister 427 genomic DNA (Cross, 1975) using the following DNA-oligonucleotide primers (KpnI and SacI restriction endonuclease recognition/cleavage sites are

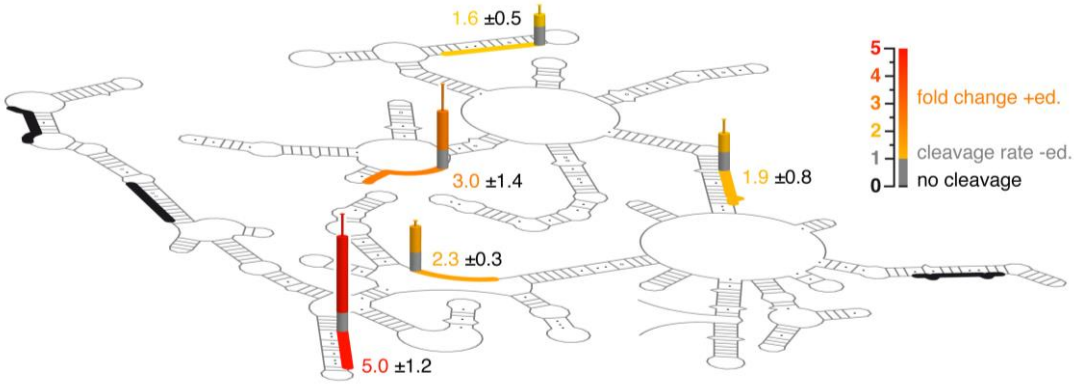


Figure II.2. Guide RNA-pre-mRNA Hybrid Formation in the Free and Editosome-Bound RNA Folding States.
Background: SHAPE-derived 2D-fold of the CYb pre-mRNA. The binding sites of eight base-complementary DNA oligonucleotides, acting as quasi gRNAs, are shown as thick lines. Colored bars indicate the change in RNaseH-based cleavage of the different DNA-pre-mRNA hybrid molecules in the free and editosome-bound RNA folding states. Gray: RNaseH cleavage in the free RNA folding state (=1). Yellow-red: >1-5-fold stimulus in the editosome-bound folding state. Black: no RNaseH cleavage. Errors are standard deviations (SD). A representative example of the data acquisition, normalization and quantification procedures is shown in Figure II.1.

underlined): CYb_forw. GGGGTACCAGCGGAG-AAAAAAGAAAGGGTC; CYb_rev. CCGAGCTCC-TAATCTAACCTACACACTATG. PCR-amplicons were cloned into the KpnI and SacI sites of phagemid pBS SK$^-$ (Invitrogen) and transcripts were generated by runoff *in vitro* transcription from linearized plasmids using T7-RNA polymerase. RNAs were purified from non-incorporated NTPs by size exclusion chromatography, EtOH-precipitated and dissolved in 10mMTris/HCl pH 7.5, 1mM EDTA (TE).

Editosome preparations

Editosomes were isolated from insect stage *Trypanosoma brucei* using the monomorphic strain Lister 427 (Cross, 1975). Cells were grown in SDM79-medium (Brun and Schönenberger, 1979) to a cell density of $1x10^7$ cells/mL. After harvesting cells were disrupted by N_2-cavitation at isotonic conditions (Hauser et al., 1996). Mitochondrial vesicles were separated in Percoll step gradients and lysed in 1% (v/v) Nonidet P-40 in editing buffer (EB: 20mM HEPES/KOH, pH 7.5, 30mM KCl, 10mM $MgCl_2$, 1mM DTT) containing 1mM PMSF, 1µg/mL leupeptin and 10µg/mL trypsin inhibitor. The detergent lysates were further separated by isokinetic ultracentrifugation in 10-35% (v/v) linear glycerol gradients as in Göringer et al., 1994 and after fractionation, 20S fractions were assayed for their RNA editing (Igo et al., 2000, Igo et al., 2002) and RNA binding activity (Katari et al., 2013). RNA binding and RNA editing-competent fractions were stored at -20°C

RNaseH-Based "gDNA" Annealing Assay.

The ability of free and editosome-bound pre-mRNAs to form gRNA-pre-mRNA hybrid RNAs was analyzed using a panel of "guiding" (g)DNA oligonucleotides complementary to different regions of the CYb-transcript (see numbers in brackets). The following oligodeoxynucleotides were used: CYb-1 (150-161): ACAAAATATCAA; CYb-2 (227-238): GTAAA-TATAATA; CYb-3 (255-266): ATTGACTTAAAT; CYb-4 (459-470): TCACTTCCCCAA; CYb-5 (651-662): AACATATCTCTC; CYb-7 (840-851): AAAA-ACAAACCC; CYb-8 (967-978): ACTCATTCATAT; CYb-9 (1045-1056): CAAAAATAATAA. Annealed gDNA-pre-mRNA hybrid molecules were identified by RNaseH cleavage. For that, co-transcriptionally [^{32}P]-labeled CYb-RNA (1nM) was incubated with an equimolar amount of editosomes for 15min at 27°C in Editing Buffer (EB: 20mM HEPES/KOH, pH 7.5, 30mM KCl, 10mM $MgCl_2$, 1mM DTT). After the addition of 10-1000nM DNA-oligonucleotide, RNaseH cleavage was performed for 20min at 27°C using 0.01U/mL *E. coli* RNaseH. RNA fragments were phenol-extracted, followed by EtOH precipitation. Cleavage products were electrophoretically separated in 8M urea-containing 5% (w/v) polyacrylamide gels, visualized by phosphorimaging and densitometrically quantified. The data processing is sketched in Figure II.1.

References

Brun, R., and Schönenberger, M. (1979). Cultivation and *in vitro* cloning or procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. Acta Trop. 36, 289-292.

Cross, G.A. (1975). Identification, purification and properties of clone specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. Parasitology 71, 393-417.

Göringer, H.U., Koslowsky, D.J., Morales, T.H., and Stuart, K. (1994). The formation of mitochondrial ribonucleoprotein complexes involving guide RNA molecules in *Trypanosoma brucei*. Proc. Natl. Acad. Sci. USA 91, 1776-1780.

Hauser, R., Pypaer, M., Häusler, T., Horn, E.K., and Schneider, A. (1996). *In vitro* import of proteins into mitochondria of *Trypanosoma brucei* and *Leishmania tarentolae*. J. Cell Sci. 109, 517-523.

Igo, R.P.Jr., Palazzo, S.S., Burgess, M.L.K., Panigrahi, A.K., and Stuart, K. (2000). Uridylate Addition and RNA Ligation Contribute to the Specificity of Kinetoplastid Insertion RNA Editing. Mol. Cell. Biol. 20, 8447-8457.

Igo R.P.Jr, Weston D.S., Ernst NL, Panigrahi A.K., Salavati R, and Stuart, K (2002). Role of uridylate-specific exoribonuclease activity in *Trypanosoma brucei* RNA editing. Eukar. Cell 1, 112-118.

Katari, V.S., van Esdonk, L. and Göringer, H.U. (2013). Molecular crowding inhibits U-insertion/ deletion RNA editing *in vitro*: consequences for the *in vivo* reaction. PLoS One 8, e83796.

Leeder, W.M., Voigt, C., Brecht, M., and Göringer, H.U. (2016). The RNA chaperone activity of the *Trypanosoma brucei* editosome raises the dynamic of bound pre-mRNAs. Sci. Rep. 6, 19309.

Missel, A., and Göringer, H.U. (1994). *Trypanosoma brucei* mitochondria contain RNA helicase activity. Nucleic Acids Res. 22, 4050-4056.

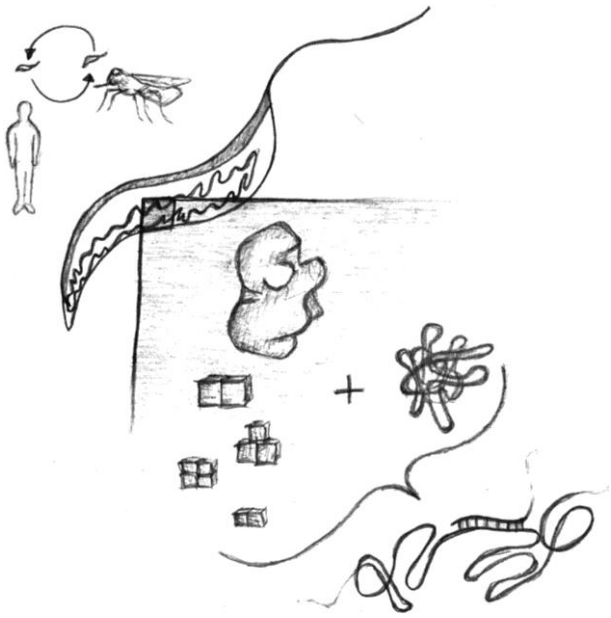
Rajkowitsch, L., Semrad, K., Mayer, O., and Schroeder, R. (2005). Assays for the RNA chaperone activity of proteins. Biochem. Soc. Trans. 33, 450-445.

Chapter III

Surface-Driven RNA Refolding by the OB-Fold Proteins of the *Trypanosoma brucei* Editosome

The following chapter has been published in:

Voigt, C., Dobrychłop, M., Kruse, E., Czerwoniec, A., Kasprzak, J.M., Bytner, P., Bujnicki, J.M., and Göringer, H.U. (2017). Surface-Driven RNA-Refolding by the OB-Fold Proteins of the *Trypanosoma brucei* Editosome. bioRxiv doi: 10.1101/099705.



Summary

RNA editing in African trypanosomes represents an RNA processing reaction that generates functional mitochondrial transcripts from sequence-deficient pre-mRNAs. The reaction is catalyzed by a macromolecular protein complex known as the editosome. Editosomes have been demonstrated to execute RNA chaperone activity to overcome the highly folded nature of pre-edited substrate mRNAs. The molecular basis of this activity is unknown. Here we test five OB-fold proteins of the editosome as potential candidates. We show that the different proteins interact by hetero-oligomerization and we demonstrate that all proteins execute RNA chaperone activity. Activity differences correlate with the surface areas of the proteins and map predominantly to the intrinsically disordered subdomains of the polypeptides. To provide a structural context for our findings we present a coarse-grained model of the editosome. The model suggests that an inner core of catalytically active editosome components is separated from an outer shell of intrinsic disordered protein domains that act as RNA remodeling sites.

Introduction

Editosomes are high molecular mass (0.8MDa, 20S) protein complexes that catalyze the U-insertion /deletion RNA editing reaction in African trypanosomes and other kinetoplastid organisms (Göringer, 2012; Aphasizhev and Aphasizheva, 2014). The processing reaction is characterized by the site-specific insertion and to a lesser extent, deletion of exclusively U-nucleotides and converts cryptic primary transcripts into translatable mRNAs. Editosomes harbor one substrate RNA binding site (Böhm et al., 2012) and interact with a large set of pre- and partially edited mitochondrial transcripts to execute the reaction. Recent structure probing experiments uncovered that the pre-edited mRNAs are unusually folded. The different transcripts have thermodynamic stabilities that resemble structural RNAs (Leeder et al., 2016a) and due to a very high content of runs of G-nucleotides they contain multiple, up to five, G-quadruplex (GQ)-folds (Leeder et al., 2016b). Perhaps as a consequence, *T. brucei* editosomes execute a complex-intrinsic RNA chaperone activity (Böhm et al., 2012). The activity acts by increasing the flexibility of predominantly U-residues to lower their base pairing probability, thereby generating a simplified RNA folding landscape with a reduced energy barrier to facilitate the binding of gRNAs (Leeder et al., 2016a, chapter II). Thus, the editosome-driven RNA unfolding reaction is important for the RNA editing cycle, especially during the initiation and elongation phases of the process. However, the molecular nature of the chaperone activity is not understood.

Proteins with RNA chaperone activity represent a structurally diverse group of polypeptides that contribute to almost all RNA-driven biochemical processes in all domains of life (reviewed in Semrad, 2011). This includes many viral and bacterial proteins such as Ncp7 (Bernacchi et al., 2002), StpA and Hfq (Zhang et al., 1995, Moll et al., 2003) and a large number of ribosomal proteins of both, pro- and eukaryotic origin (Semrad et al., 2004). Importantly, several of the RNA chaperones contain one or more OB-fold motif(s). OB refers to the general oligonucleotide/oligosaccharide binding ability of the proteins, which is mediated by a five-stranded β-sheet fold that is coiled up to form a closed β-barrel (Murzin, 1993). OB-folds have specifically been identified in bacterial and plant cold shock proteins (Csp), where they contribute to resolve misfolded RNA species (Bae et al., 2000; Nakaminami et al., 2006). Furthermore, OB-fold proteins are universally involved in RNA remodeling steps prior to the initiation of protein biosynthesis with ribosomal protein S1 and *E. coli* initiation factor 1 (IF1) containing multiple OB-domains (Duval et al., 2013; Croitoru et al., 2006).

The protein inventory of the *Trypanosoma brucei* 20S editosome lists six OB-fold proteins termed TbMP81, TbMP63, TbMP42, TbMP24, TbMP19 and TbMP18. TbMP stands for *Trypanosoma brucei* mitochondrial protein, followed by a number indicating the calculated molecular mass in kDa (Worthey et al., 2003). TbMP81, TbMP63 and TbMP42 additionally contain two C2H2-type Zn-fingers or C2H2-Zn-finger-like domains and all six proteins have recently been predicted to harbor long stretches of intrinsically disordered regions (IDR) (Czerwoniec et al., 2015). The domain structure of the different proteins, a 3D-consensus model of the OB-fold and a plot of their disorder propensities (Kozlowski and Bujnicki, 2012) is summarized in Figure III.1. While the percentage of predicted disorder varies between 25% in TbMP18 and 72% in TbMP81, together, the six proteins are more disordered than the average of all other editosomal proteins (Czerwoniec et al., 2015). Two of the proteins (TbMP63, TbMP81) even contain more disordered than ordered regions.

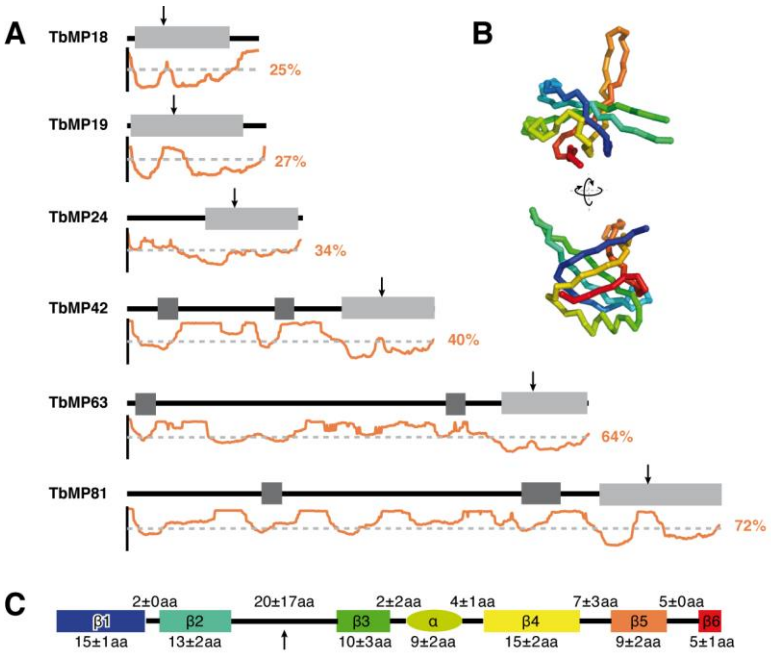


Figure III.1. Domain Structure of the OB-Fold Proteins of the T. brucei Editosome.
A. OB-fold domains: boxes in light gray. C2H2-Zn-fingers and C2H2-Zn-finger-like domains: boxes in dark gray. Arrow: Position of the loop (L)-sequence between β-sheets β2 and β3 (L_{23}). Disorder propensity plots (between 0 and 1) for each protein are shown in orange. The gray dashed line indicates a value of 0.5. Contents of disorder are listed as percent values.
B. Consensus structure of all editosomal OB-folds in spectral colors from the N-terminus in blue to the C-terminus in red. For details see Table S1.
C. Sketch of the secondary structural elements (colored as in B) of the consensus OB-fold structure. Rectangles: β-sheets. Oval sphere: α-helix. Black lines: loop regions. Numbers: average amino acid length (±SD). Details are given in Table S2.

Although several of the editosomal OB-fold proteins have been investigated in gene ablation experiments and as recombinant proteins, no function for any of the six polypeptides has emerged from these studies. TbMP24, TbMP18 and TbMP42 were identified as RNA binding proteins (Tarun et al., 2008; Salavati et al., 2006; Brecht et al., 2005). Recombinant TbMP42 was characterized to execute endo/exonuclease activity (Brecht et al., 2005; Niemann et al., 2008) and recombinant TbMP24 was shown to perform RNA annealing activity (Kala and Salavati, 2010). Gene knockdown of five of the proteins (TbMP18, TbMP24, TbMP42, TbMP63, TbMP81) identified an impact on the structural integrity of the editosome (Drozdz et al., 2002; Huang et al., 2002; Guo et al., 2008; Salavati et al., 2006; Law et al., 2007), which was supported by yeast two-hybrid experiments in combination with protein-protein interaction studies (Schnaufer et al., 2010). The data suggest a scenario in which TbMP18, TbMP24, TbMP42, TbMP63 and TbMP81 interact with each other, mainly relying on their OB-fold domains thereby forming a clustered "OB-fold core" within the 20S editosome (Park et al., 2012). Recent inter-protein chemical cross-linking data support the existence of such a "core domain" and have identified cross-links between all six OB-fold proteins, most of which within or in proximity to the OB-fold domains (McDermott et al., 2016).

Importantly, clustered OB-folds in the yeast RRP44 protein have been demonstrated to catalyze the unwinding of dsRNA (Lorentzen et al., 2008; Bonneau et al., 2009). As a consequence, Böhm et al., 2012 suggested that the potential OB-fold core of the *T. brucei* editosome might act in a similar fashion. Here we show that the *T. brucei* editosomal OB-fold proteins indeed catalyze an RNA refolding reaction. We demonstrate that the activity is inherent to the intrinsically disordered protein (IDP)-regions of the different proteins and we uncovered a correlation of the chaperone activity to the surface areas of the proteins. Using a coarse-grained modeling approach, we provide a molecular model of the *T. brucei* 20S editosome, which suggests that the high molecular mass complex has a bi-partite composition: an outer shell of intrinsically disordered proteins that act as RNA contact and RNA remodeling elements and an inner core of structurally defined proteins, mediating the catalytic reactions of the complex.

Results

Recombinant Expression of the OB-Fold Proteins of the *T. brucei* Editosome

To test whether the editosome-inherent RNA chaperone activity is mediated by one or several of the editosomal OB-fold proteins (Böhm et al., 2012) we expressed the six proteins as recombinant polypeptides in *E. coli*. TbMP81, TbMP63, TbMP42 and TbMP24 were expressed as full-length (FL) and as OB-fold-only (OB) constructs. TbMP19 and TbMP18, due to their small size, were expressed as FL-constructs only. While the different constructs expressed well upon induction, the majority of polypeptides remained insoluble at an expression temperature of 37°C. Lowering the temperature to 18°C improved the solubility, however, all attempts to express soluble TbMP19-FL and TbMP63-FL failed. For TbMP63-OB and TbMP24-FL we devised a co-expression regime together with TbMP18-FL as a previously identified interaction partner (Schnaufer et al., 2010). In the end, 8 of the 10 recombinant protein constructs were available in yields between 5-40mg soluble protein/L *E. coli* culture (Figures III.2A and Table S4).

Notably, all expressed protein constructs formed multiple homo-oligomeric higher order assemblies with one dominant assembly state (Figure III.2B). TbMP81-FL and TbMP81-OB predominantly formed homodimers, TbMP42-FL, TbMP42-OB and TbMP24-OB homotrimers and TbMP18-FL homotetramers. Importantly, the identified oligomers of the FL and OB-only constructs were invariably identical, indicating that the interaction surfaces involve the OB-fold domains of the different proteins. In the case of the two co-expressed constructs (TbMP24-FL/TbMP18-FL, TbMP63-OB/TbMP18-FL) we exclusively identified hetero-oligomeric assemblies (Figure III.2B), suggesting an enhanced stability and/or solubility of the hetero-oligomeric complexes over the homo-oligomeric complexes.

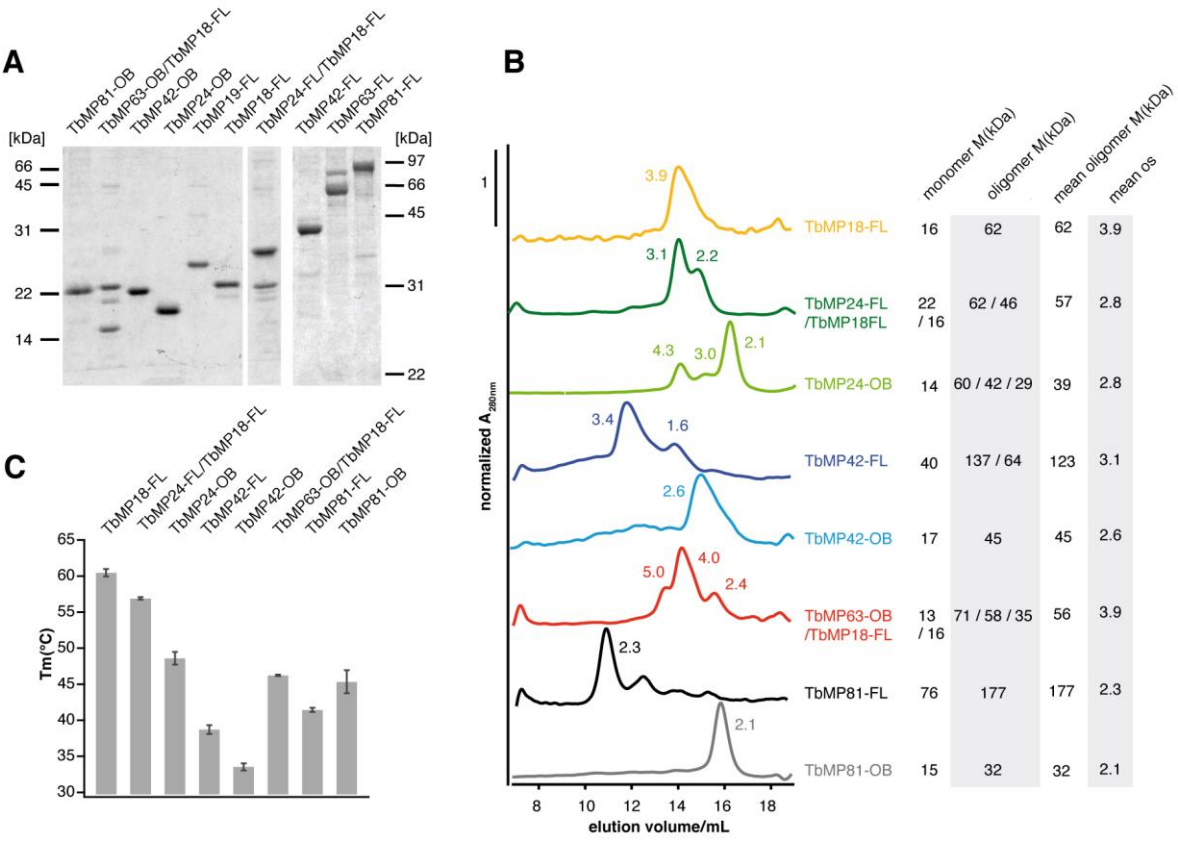
Structural Characteristics of the Recombinant OB-Fold Proteins

The purified protein preparations were analyzed for their secondary (2D)-structure content, using circular dichroism (CD)-spectroscopy (Figure III.2D). As expected and in line with structure prediction algorithms, all proteins display a high fraction of unstructured sequence stretches. The α-helical content of the different proteins is as low as 6-17%, the β-sheet content varies between 25-38% and the amount of coil structures varies between 45-65%. On average 33% of the proteins are disordered. To further assess the characteristics of the proteins, we measured their thermal unfolding, using differential scanning fluorimetry (Niesen et al., 2007). Intrinsically disordered proteins typically display broad denaturation profiles and low cooperativity in the folding /unfolding transition (Kazakov et al., 2009). All melting curves were converted into fraction-folded (α) *vs.* temperature plots to derive half-maximal melting temperatures (T_m) (Figure S1). The data are summarized in Figure III.2C. Despite the fact that all proteins contain a similarly structured OB-fold, they cover a broad thermal stability range: The OB-fold-only construct of TbMP42 is the least stable polypeptide (T_m 34°C) and TbMP18-FL represents the most stable protein (T_m 61°C). Furthermore, by comparing the FL and OB-fold-only versions of the same proteins, we uncovered that OB-folds can be stabilized as well as destabilized by the remaining amino acid sequences. TbMP42-OB is stabilized with a ΔT_m of 5.2°C and TbMP81-OB is destabilized with a ΔT_m of 3.9°C. Thus, sequences outside the OB-folds modulate the thermal stability of the different protein

constructs. The same holds true for the bimolecular complexes TbMP18-FL/TbMP24-FL and TbMP18-FL/TbMP63-OB. Both complexes have a lower T_m than TbMP18-FL alone.

The Formation of Hetero-Oligomeric Complexes

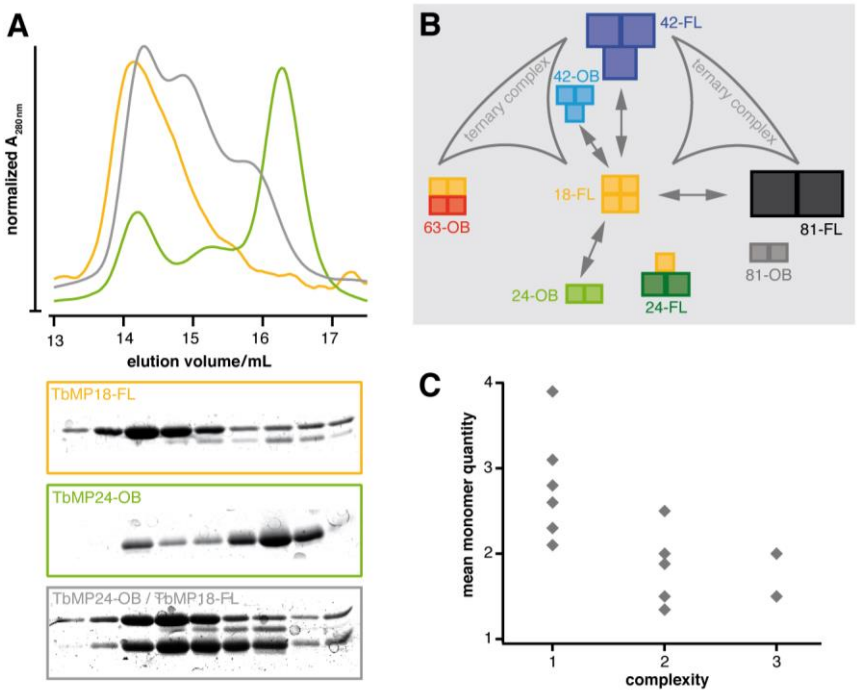
Editosome subcomplex reconstitution experiments, as well as recent chemical cross-linking experiments, demonstrated that the OB-fold proteins of the *T. brucei* editosome interact with each other (Schnaufer et al., 2010; Park et al., 2012; McDermott et al., 2016). This tempted us to analyze the association behavior of the different protein preparations in "mix-and-match-type" binding experiments. Figure III.3A shows, as a representative example, the formation of a trimeric (1:2) TbMP18-FL/TbMP24-OB complex. Pairwise interaction experiments were conducted with all combinations of proteins and we identified hetero-oligomers for TbMP18-FL with TbMP24-OB, TbMP42-FL, TbMP42-OB and TbMP81-FL (Figures III.3B, III.3D). Importantly, no pairwise interaction was detected without TbMP18-FL. Furthermore, TbMP42-FL, as well as TbMP42-OB, was able to



D

	prediction(%)				CD spectra(%)			
	helix	sheet	coil	disordered	helix	sheet	coil	disordered
TbMP18-FL	13	40	46	25	17	38	45	20
TbMP24-FL/TbMP18-FL	9	36	55	29	6	35	59	39
TbMP24-OB	4	43	52	16	8	34	58	37
TbMP42-FL	6	26	69	40	6	38	55	34
TbMP42-OB	12	44	44	12	8	34	58	34
TbMP63-OB/TbMP18-FL	12	45	42	18	15	30	55	32
TbMP81-FL	17	12	71	72	10	25	65	42
TbMP81-OB	12	43	45	29	9	35	55	28

Figure III.2. Characterization of Recombinant Editosomal OB-fold Proteins.
A. Gel electrophoretic analysis of purified protein preparations in 15% (w/v) (left) and 10% (w/v) (right) SDS-containing polyacrylamide gels. Apparent molecular masses are in kDa.
B. Normalized size exclusion chromatography profiles of purified OB-fold proteins. Peak numbers indicate oligomerization states (os), which were determined as outlined in the experimental procedures section. The molar ratios of the co-expressed protein constructs were determined as 2:1 and 1.5:1 for the TbMP24-FL/TbMP18-FL construct and as 3:2, 2:2 and 1:1 for the TbMP63-OB/TbMP18-FL complex.
C. Differential scanning fluorimetry (DSF)-based melting transitions (T_m) of the different protein constructs. Error bars are SDs. For an example see Figure S1.
D. Circular dichroism (CD)-derived secondary structure content of recombinant OB-fold proteins. Predicted secondary structure contents were derived from the homology models (see Table S1 and Table S2). The structure of non-modeled protein domains was predicted using PSIPRED (McGuffin et al. 2000). The molar ratios of the co-expressed constructs were determined as 1.1:1 for the TbMP24-FL/TbMP18-FL complex and as 1:1.5 for the TbMP63-OB-TbMP18-FL complex (see also Table S4). Disorder propensities were calculated using MetaDisorderMD2 (Kozlowski and Bujnicki 2012).



D

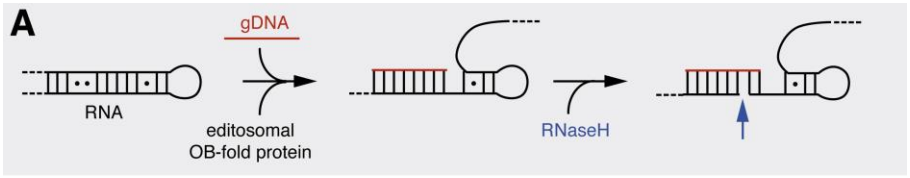
	M (kDa)	molar ratio	os
MP18-FL / MP24-OB	50	1 : 2	3
MP18-FL / MP42-FL	140	n.d.	4 - 6
MP18-FL / MP42-OB	80	2 : 2	4
MP18-FL / MP81-FL	180	n.d.	3 - 5
MP18-FL / MP42-FL / MP81-FL	300	2 : 2 : 2	6
MP18-FL / MP42-FL / MP63-OB	110	n.d.	4 - 5

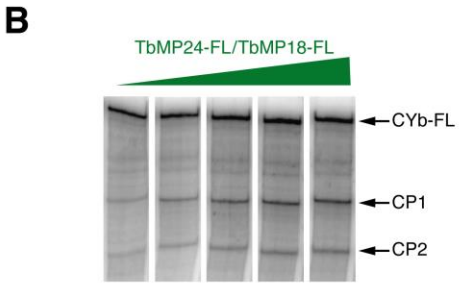
Figure III.3. The Formation of OB-Fold Protein Complexes *in vitro.*
A. Normalized size exclusion chromatography profiles (upper panel) and gel electrophoretic analysis (lower panels) of a representative experiment to demonstrate the formation of a hetero-oligomeric TbMP24-OB-TbMP18-FL complex (gray). Green: TbMP24-OB. Yellow: TbMP18-FL. See also Figure S2.
B. *In vitro* interaction map. Summary of all identified protein interactions (arrows). The size of the colored boxes indicates the apparent molecular sizes of the different proteins and the number of boxes refers to the number of monomeric units in the most abundant complex. Curved triangles: ternary interactions.
C. Correlation of the mean number of protein monomers in all homo- and hetero-oligomeric complexes as a function of the complexity of the complexes, with respect to the number of unique subunits.
D. Summary of the characteristics of the hetero-oligomeric complexes. os: oligomerization state. n.d.: not determined.

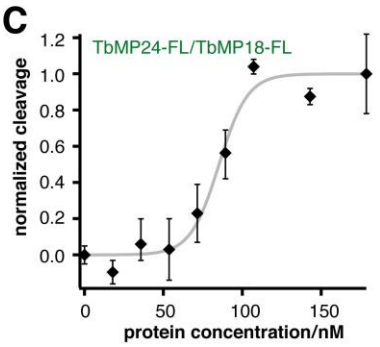
hetero-oligomerize with TbMP18-FL, again indicating that the TbMP42-OB-fold is sufficient to mediate the interaction. The same holds true for the association of TbMP24 with TbMP18. Thus, the individual OB-folds likely act as docking modules.
We also identified two ternary complexes: First, a complex between TbMP42-FL, TbMP18-FL and TbMP63-OB, which was identified before (Schnaufer et al., 2010) and second, a ternary assembly between TbMP81-FL, TbMP42-FL and TbMP18-FL (Figure S2). The two ternary complexes are mutually exclusive, since we were not able to identify a complex containing both, TbMP81-FL and TbMP63-OB. Also, no ternary complex containing TbMP24 was found. Importantly, all heteromeric complexes contain less identical monomer subunits than the homomeric complexes (Figure III.3C). Thus, the proteins interact by re-organizing the single monomers instead of adding up the different homo-oligomers to larger heteromeric assemblies. A summary of the data is shown in Figures III.3B, 3D.

The OB-Fold Proteins of the *T. brucei* Editosome Execute RNA Remodeling Activity

To quantitatively assess the potential RNA remodeling activity of the different OB-fold proteins, we relied on the recently described RNaseH-based guide (g)DNA-annealing assay (Leeder et al., 2016a, chapter II). The assay is based on the rationale that a remodeled *i.e.* structurally open target RNA should bind a guide RNA-mimicking, short complementary DNA oligonucleotide more readily when compared to a structurally constrained RNA. The formed pre-mRNA-DNA oligonucleotide hybrid molecules were identified by RNaseH cleavage followed by an electrophoretic separation of the resulting RNA fragments (Figure III.4A). As a representative *T. brucei* mitochondrial transcript we used the pre-edited mRNA of apocytochrome b (CYb). The RNA chaperone activity of 20S editosomes served as a positive control (Böhm et al., 2012; Leeder et al., 2016a, chapter II). Importantly, the experiments were performed with varying amounts of protein to use the protein concentration required to achieve half-maximal cleavage ($c_{1/2}$) as a metric for the RNA remodeling activity.
All purified OB-fold homo-oligomers, co-expressed hetero-oligomers, as well as the *in vitro* formed TbMP24-OB/TbMP18-FL hetero-oligomer, were tested. Figures III.4B, III.4C show a representative example for the TbMP24-FL/TbMP18-FL complex. All data are summarized in Figure III.4E. Surprisingly, all protein constructs show RNA remodeling activity. The measured $c_{1/2}$-values range from 57±13nM (TbMP81-FL) to 164±16nM (TbMP81-OB) with a mean of 108±12nM. This represents a roughly twentyfold higher protein concentration in comparison to 20S editosomes ($c_{1/2}$ 5±0.4nM). Thus, although all protein constructs are capable of remodeling the pre-mRNA, they are less active than 20S editosomes. This suggests that the RNA remodeling activity of the editosome likely represents a cumulative trait, to which each of the







D

	$c_{1/2}$(nM)	norm. surface
TbMP18-FL	109±8.0	1.6
TbMP24-FL/TbMP18-FL	79±3.3	1.5
TbMP24-OB	111±11.5	1.1
TbMP24-OB/TbMP18-FL	143±31.3	1.3
TbMP42-FL	78±10.4	2.5
TbMP42-OB	121±11.3	1.3
TbMP63-OB/TbMP18-FL	109±3.9	1.5
TbMP81-FL	57±13.3	3.1
TbMP81-OB	164±19.5	1.0
20S editosomes	5±0.4	8.5

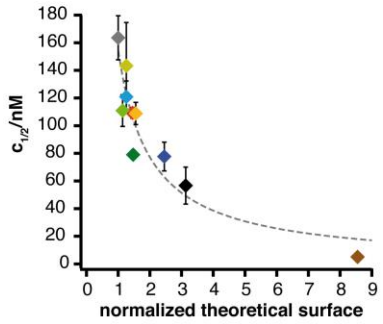


Figure III.4. RNA Chaperone Activity of the Editosomal OB-Fold Proteins.
A. Sketch of the RNaseH-based gDNA-annealing assay.
B. Representative gel electrophoretic separation of RNaseH-induced RNA cleavage products (CP1/CP2) of pre-edited CYb-mRNA, using increasing concentrations of TbMP24-FL/TbMP18-FL: 0, 70, 110, 140, 210nM (left to right). CYb-FL: full length CYb-pre-mRNA.
C. Representative plot of normalized cleavage values as a function of the TbMP24-FL/TbMP18-FL concentration. Average values with error bars indicating the maximal deviation. Gray line: Sigmoidal fit of the data to derive half-maximal cleavage concentrations ($c_{1/2}$) as a metric for the RNA chaperone activity.
D. Summary of all $c_{1/2}$-values (±SD) and normalized surface areas of the tested editosomal OB-fold constructs in comparison to 20S editosomes.
E. Plot of the derived $c_{1/2}$-values of all tested OB-fold complexes in relation to their normalized surface areas. Dotted line in gray: Reciprocal fit of the data. Error bars are SD-values

different OB-fold proteins contributes a defined increment. Furthermore, we noticed that all FL-proteins are more active than the OB-fold-only constructs. This indicates that protein sequences outside the OB-fold domains, *i.e.* the intrinsically disordered regions of the proteins, are for the most part responsible for the remodeling activity.
Importantly, for intrinsically disordered RNA binding proteins it was shown that the size of the RNA binding surfaces correlate with the overall sizes of the proteins (Wu et al., 2015). As a consequence, we analyzed whether the RNA remodeling activities of the different OB-fold proteins can be correlated to their surface areas. For that, we calculated relative surface areas for all proteins and plotted them in relation to the measured RNA remodeling activities ($c_{1/2}$-values). Figure III.4D shows the resulting plot. The two variables show a reciprocal dependency and the data fit to a hyperbole with a correlation coefficient (r^2) of 0.87. Thus, a major determinant of the RNA chaperone activity of the different OB-fold proteins is their surface area. Proteins with larger surfaces such as TbMP81-FL execute a higher refolding activity and *vice versa*. The high remodeling activity of the editosome ($c_{1/2}$ 5±0.4nM) is the result of the large surface of the 0.8MDa complex.

Structural Characterization of the RNA Remodeling Activity

As a follow up of the results above, we asked the question how the RNA remodeling activities of the different OB-fold proteins manifest on a structural level and how they compare to the structural changes induced by 20S editosomes (Leeder et al., 2016a). For that, we mapped the chaperone-induced structure changes with nucleotide resolution using SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) chemical probing (Merino et al., 2005; Low and Weeks, 2010). TbMP18-FL was selected as a representative OB-fold protein and the pre-edited mRNA of RPS12 as a typical *T. brucei* mitochondrial transcript. Figure III.5A shows the SHAPE-reactivity profile of the RNA in the presence of TbMP18-FL. As anticipated, the protein induces an 18% increase in the mean SHAPE-reactivity (Figure III.5B), which is in line with a remodeling reaction similar to 20S editosomes (26%). Twenty two percent of the nucleotide positions in the RPS12-transcript are responsive to the refolding reaction and as evidenced in the difference (Δ) SHAPE-profile (Figures III.5C, III.5D), the majority of affected nucleotides (71%) increase in flexibility. Again, this represents a value comparable to the remodeling activity of the 20S editosome (72%).

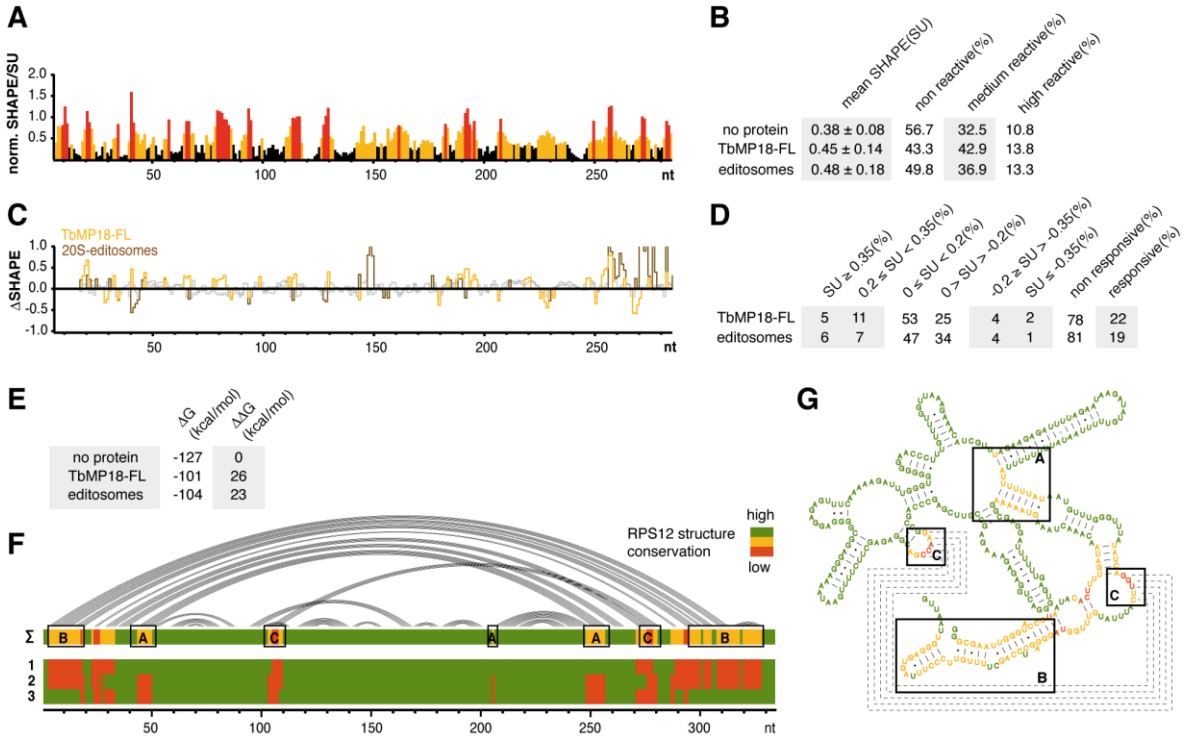


Figure III.5. SHAPE-Chemical Modification of RPS12 pre-mRNA in the Presence/Absence of TbMP18-FL.
A. Normalized SHAPE-reactivity profile of RPS12 pre-mRNA in the presence of TbMP18-FL. Black: non-reactive (SU<0.35), yellow: medium reactive (0.35≤SU<0.8), red: highly reactive (SU≤0.8) nucleotide positions.
B. Statistical summary of SHAPE-reactivity data. Mean SHAPE-reactivities were calculated from minimally three experiments and are listed ±SD. SU: SHAPE-unit
C. Difference (Δ)SHAPE-reactivity plots: Yellow: plus/minus TbMP18-FL. Brown: plus/minus 20S editosomes. Non-responsive nt-positions are in gray.
D. Statistical summary of the ΔSHAPE-reactivities.
E. Calculated Gibbs free energies (ΔG) and ΔΔG-values of RPS12 pre-mRNA in the presence/absence of TbMP18-FL and 20S editosomes using the SHAPE-reactivities as pseudo free energy constraints.
F. Arc representation of the base pairing pattern of RPS12-RNA in its free conformation. Pairwise comparison of the structure conservation from high (green) to medium (yellow) to low (red). Row 1: free RNA/+editosomes, row 2: free RNA/+TbMP18-FL. Row 3: +editosomes/+TbMP18-FL. Σ: cumulative score to demonstrate the conservation in all three structures. Boxed areas (A, B, C) indicate sequence domains with structural differences.
G. Secondary structure of free RPS12 pre-mRNA with colors indicating the structure conservation as in (F), emphasizing boxed areas A, B, C. Stippled lines: RNA-pseudoknot.

Using the SHAPE-reactivities as pseudo free energy values, we calculated minimal free energy (MFE)-structures for the RPS12-RNA in both conformational states (Figure III.5E). As a free RNA, the transcript is characterized by a Gibbs free energy (ΔG) of -127kcal/mol. In the presence of TbMP18-FL, the RNA adopts a fold of only -101kcal/mol. This demonstrates a destabilization of the RNA with a ΔΔG of -26kcal/mol. A comparison of all probed structures (free RPS12-RNA, TbMP18-FL-bound RNA and 20S editosome-bound RNA) identified that about 70% of the structural details are shared between the three RNA folds (Figure III.5F). Differences map to three regions (A, B, C in Figures III.5F, III.5G), involving a pseudoknot and the two termini of the RNA. Together, the SHAPE-chemical probing data corroborate the results of the RNaseH-based gDNA annealing assay and demonstrate that TbMP18-FL refolds and destabilizes the RPS12-transcript with qualitative and quantitative characteristics similar to 20S editosomes.

Modeling the RNA Refolding Domain(s) of the 20S Editosome

Lastly, we asked the question whether the collected data, especially the OB-protein interaction data can be used to derive a structural model of the RNA refolding domains of the *T. brucei* 20S editosome. For that, we performed computational modeling using the program PyRy3D (http://genesilico.pi/pyry3d/), specifically designed to calculate low-resolution models of high molecular mass complexes that contain intrinsic disorder. The program allows the usage of experimentally as well as computationally-derived atomic coordinates together with flexible shapes to include disordered substructures. It has been used to model the structures of several proteins and complexes involved in RNA metabolism, including the CCR4-Not complex (Ukleja et al., 2016). PyRy3D simulations rely on spatial restraints to implement experimentally-derived interaction data and a scoring function fits the individual components into the contour map of the complex. For that, we used the cryo-EM structure of the *T. brucei* 20S editosome (Golas et al., 2009). As spatial restraints we implemented all above described

OB-fold interaction data, as well as all published binary interactions of editosome components, including the recent chemical cross-linking data of the *T. brucei* OB-fold proteins (Schnaufer et al., 2010; Park et al., 2012; McDermott et al., 2016). Structure coordinates for the individual proteins were taken from Czerwoniec et al., 2015 (and references therein) and all disordered regions were simulated as coarse-grained flexible shapes. Three hundred independent simulations (300000 steps each) were performed and the resulting 300 models were clustered as outlined in the Supplemental Experimental Procedures. Figures III.6A and III.6B show the medoid structure of the largest cluster (54 models). The different proteins fit tightly into the available volume of the 20S editosome EM-density map, with a cross-correlation coefficient of 0.64. A dissection of the structurally well-defined parts of the model from the positions of the intrinsically disordered domains of the different proteins is shown in Figure III.7A,B. This identifies that structured protein domains are preferentially located inside the 0.8MDa complex, while all IDP-domains seem to be preferentially located in peripheral regions of the particle (Figure III.7C). This suggests a general bi-partite domain composition of the editosome: an outer shell of intrinsically disordered proteins, which act as RNA contact and RNA remodeling elements and an inner core of structurally well-defined proteins, which mediate the catalytic reactions of the complex.

Discussion

Editosomes execute a complex-inherent RNA chaperone activity to remodel the highly folded structures of mitochondrial substrate pre-mRNAs (Böhm et al., 2012; Leeder et al., 2016a, chapter II). The activity has been characterized to simplify the folding landscape of the different RNAs to facilitate the annealing of gRNAs as templates in the reaction. While several protein components of the editosome have been identified to catalyze defined steps of the editing cycle (reviewed in Göringer, 2012), no editosomal protein has as of yet been recognized to mediate the RNA chaperone function.

Here, we provide evidence that the OB-fold proteins of the editosome execute RNA remodeling activity. For that, we generated recombinant versions of the different proteins and identified that they are highly unstructured. Furthermore, our CD-measurements support computational predictions suggesting that the OB-fold proteins of the *T. brucei* editosome are in large parts intrinsically disordered (Czerwoniec et al., 2015). The disordered regions not only locate outside the OB-fold domains, they can also be found in the loop regions of the different OB-folds and here specifically in the loop sequences between β-sheets β2 and β3 (L_{23}) (Figure III.1C). We also confirmed previous observations (Schnaufer et al., 2010; Park et al., 2012; McDermott et al., 2016) that the different proteins have a high propensity to homo- and hetero-oligomerize, using their OB-folds as interaction domains. Interestingly, all identified pair-wise inter-

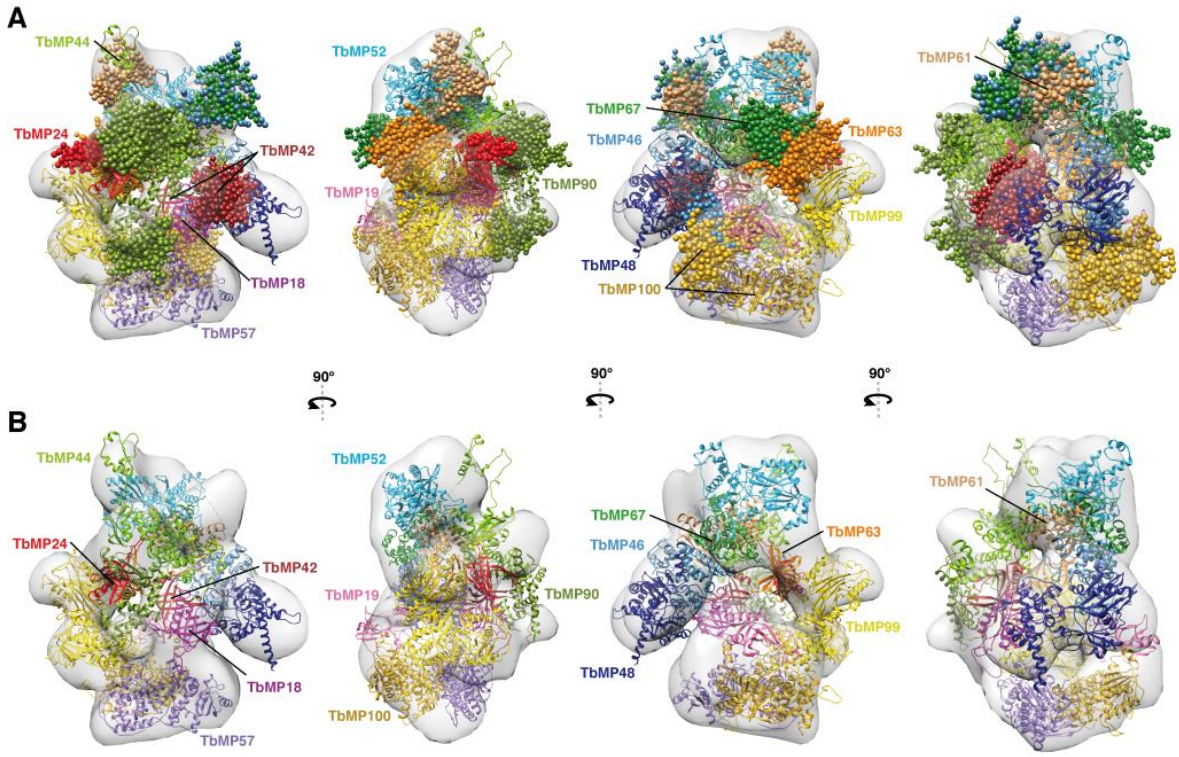


Figure III.6. Structure Model of the *T. brucei* 20S Editosome.
A. Rotational images (90°) of the PyRy3D-derived 20S editosome model based on the cryo-EM density map of Golas et al., 2009. The cryo-EM map is shown as a gray surface in the background at a contour level of 0.0419. Individual proteins are color -coded and labeled. All distance restraints used in the calculation are listed in Table S5. PyRy3D can be accessed at (http://genesilico.pi/pyry3d/).
B. 20S editosome model as in (A), omitting the intrinsically disordered protein domains for clarity.

actions involve TbMP18, the smallest of the editosomal OB-fold proteins, possibly implicating a key role in the assembly of the editosomal complex.
Importantly, all homo- and hetero-oligomeric complexes showed RNA chaperone activity. The activities correlate with the surface size of the different proteins, indicating that the process is primarily surface-driven with perhaps cumulative characteristics: every OB-fold protein provides a defined value of RNA unfolding activity, which together generates the overall activity of the editosome. This is supported by the fact that editing-active editosome preparations showed the highest activity. In addition, all FL-constructs displayed higher refolding activities in comparison to the OB-fold-only polypeptides, suggesting that the chaperone activity is primarily located within the disordered regions of the proteins. This represents a different situation to the yeast RRP44 protein, where the RNA unwinding reaction is mediated by a cluster of three OB-folds (Lorentzen et al., 2008; Bonneau et al., 2009). However, as anticipated, the refolding reaction induces a destabilization of the substrate RNAs, thereby generating RNA minimal free energy structures of reduced thermodynamic stability. As such, the individual OB-fold proteins execute a refolding reaction that is biochemically and structurally equivalent to the activity of the editosomal complex.
As a consequence of the described results, we propose a scenario in which the OB-folds and the IDP-sequences of the different proteins execute separate but interdependent functions: The OB-folds primarily act as docking modules to assemble an "OB-core" within the editosome as previously discussed (Park et al., 2012), which positions the IDP-sequences of the different proteins on the surface of the catalytic complex, where they function to bind and refold mitochondrial pre-mRNAs. Thus, our data add to the

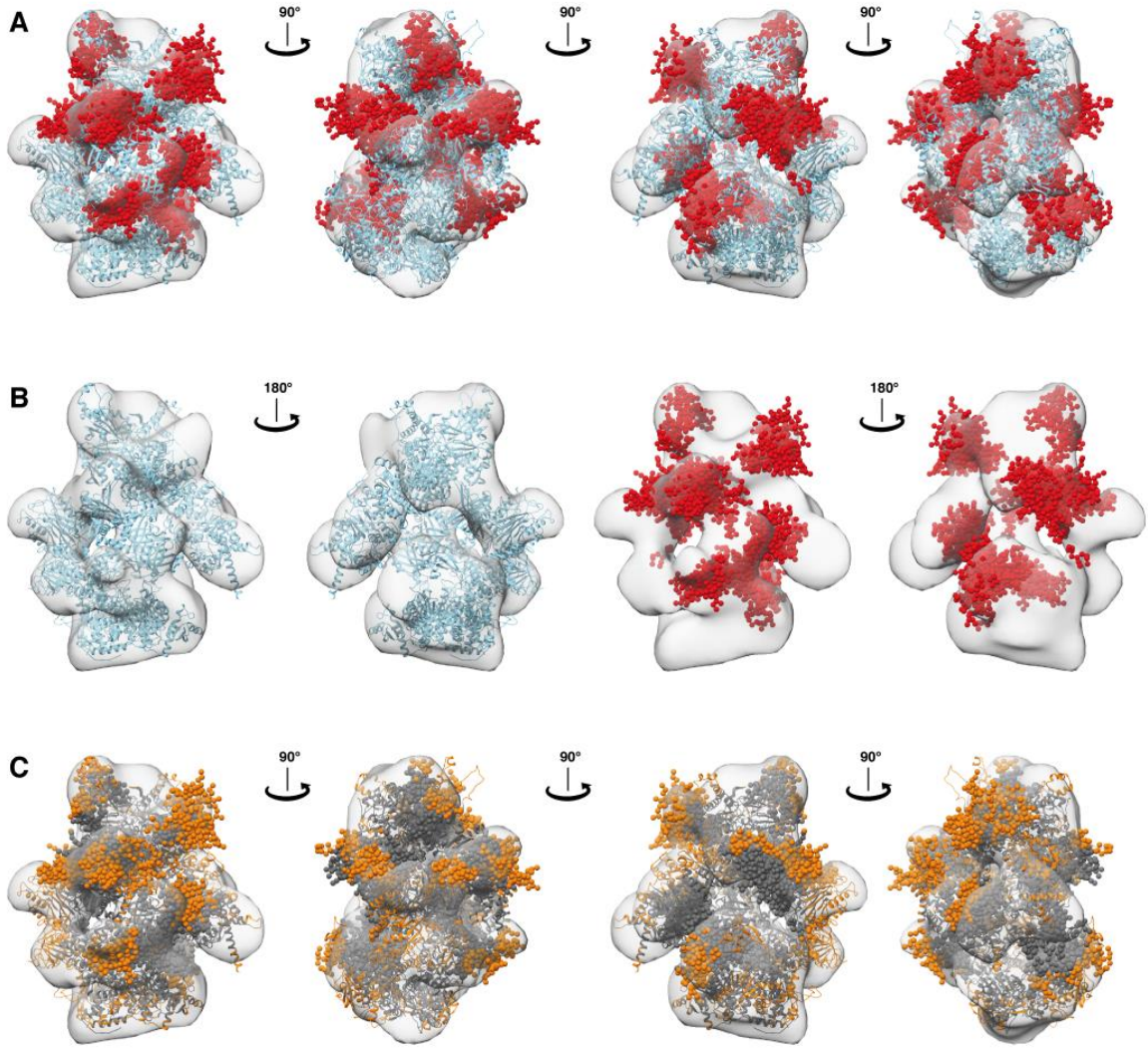


Figure III.7. Order/Disorder Distribution in the 20S Editosome.
A. Rotational images (90°) of the PyRy3D-derived 20S editosome model specifically emphasizing the intrinsically disordered regions as clustered spheres in red and structurally ordered regions in light blue. The cryo-EM-based shape of the complex (Golas et al., 2009) is shown as a gray surface (contour level 0.0419).
B. Rotational images (180°) of the editosome model, showing only the ordered (light blue) or disordered (red) regions of the complex. Ordered protein domains have a surface/core-ratio of 30/70. IDP-domains are characterized by a value of 40/60.
C. Rotational images (90°) of the editosome model, emphasizing surface-located protein domains in orange and structurally "buried" residues in gray.

functional assignment of the protein inventory of the editosome and within this context it is important to note that about half of the 12 core proteins of the editosome are allocated to remodel RNA. This indicates the importance of the process within the editing reaction cycle. Of course, we cannot exclude that IDP-sequences of other, non-OB-fold editosomal proteins such as TbMP90, TbMP67, TbMP61 or TbMP100 might contribute as well (Czerwoniec et al., 2015).

Intrinsic disorder is a common feature of many RNA binding proteins (Tompa and Csermely, 2004; Varadi et al., 2015). It provides the functional advantage of allowing multiple contacts during the initial binding reaction, while at the same time enabling sufficient conformational flexibility to target different RNA ligands (Varadi et al., 2015). In addition, intrinsic disorder has a kinetic advantage since an increased capture radius results in a higher binding probability. This mechanism is known as "fly-casting mechanism" (Shoemaker et al., 2000). It accounts for the fact that structurally flexible protein sequences can enlarge their effective "reach", which is of special importance for membrane-bound macromolecular complexes, especially in crowded solvent conditions. RNA editing takes place in the highly crowded mitochondria of African trypanosomes and preliminary evidence for an attachment of the editosome to mitochondrial ribosomes and/or the mitochondrial inner membrane have been discussed (Aphasizheva et al., 2011; Katari et al., 2013). We propose that the energy for the RNA refolding reaction, carried out by the editosome, is provided by an "entropy transfer"-mechanism (Tompa and Csermely, 2004). The binding of a substrate RNA induces a structural re-orientation of the OB-fold proteins, which results in a loss of conformational entropy. This entropy is used to remodel the bound RNA in agreement with the observation that the reaction does not require ATP (Böhm et al., 2012). Our data further suggest that the pre-edited substrate mRNAs likely interact with the editosome by two successive RNA binding modes. As in the case of the CBP2-bI5 group I intron interaction (Bokinsky et al., 2006), an initial, rather non-specific, interaction induces a set of conformational fluctuations in the bound RNA, which is followed by a slow, specific binding mode that stabilizes the processing-competent RNA conformation. Support for such a scenario comes from the observation that flexibility restrictions in model editing substrate RNAs inhibit RNA editing *in vitro* (Leeder et al., 2015).

As a consequence of our hypothesis, we postulate that the IDP-domains of the different OB-fold proteins are located on the surface of the 20S editosome. To provide a structural context for the assumption, we performed a coarse-grained modeling of the structure of the *T. brucei* editosome using published cryo-EM data (Golas et al., 2009) and all OB-fold interaction data as spatial restraints (Schnaufer et al., 2010; Park et al., 2012; McDermott et al., 2016). Interestingly, while the model agrees with the anticipated surface location of the different IDP-domains in two regions of the editosome, it also suggests a clustering of all structurally well-defined proteins in the center of the complex. Thus, the complex separates an inner core of catalytically active editosome components from an outer shell of ID-proteins that act as RNA contact and RNA remodeling elements. Interestingly, a similar separation of structurally defined protein regions from intrinsically disordered domains was also identified for the human spliceosome (Korneta and Bujnicki 2012). Thus, the described organization might reflect a more general structural scenario of macromolecular ribonucleoprotein machineries.

Author contributions

H.U.G. conceived and supervised the project. C.V. and E.K. conducted the experiments. J.M.B. supervised the computational analysis of the editosome structure. M.D., J.M.K., and A.C. carried out the modeling. P.B. carried out computational analyses of intrinsic disorder. All authors contributed to the analysis and interpretation of the data. H.U.G. and C.V. wrote the manuscript with input from all authors.

Acknowledgements

We thank Matthias Leeder for his help conducting the SHAPE-experiments and Wim Hol (University of Washington, Seattle) for plasmid constructs. This work was supported by the German Research Foundation (DFG-SFB902) and the Dr. Illing Foundation for Molecular Chemistry to H.U.G. The modeling work was supported by the Polish Ministry of Science and Higher Education (grants 0083/IP1/2011/71 and N N301 123138) and by the European Research Council (StG grant RNA+P=123D to J.M.B.). M.D. was supported by the KNOW RNA Research Centre in Poznan (grant 01/KNOW2/2014). J.M.B. was also supported by the 'Ideas for Poland' fellowship from the Foundation for Polish Science.

Experimental procedures

Full experimental procedures are provided in the Supplemental Experimental Procedures.

Expression and Purification of Recombinant OB-Fold Proteins

DNA sequences of the six *T. brucei* editosomal OB-fold proteins (TbMP81, TbMP63, TbMP42, TbMP24, TbMP19, TbMP18) were PCR-amplified from *T. brucei* genomic DNA (strain Lister427) (Cross 1975), using the DNA oligonucleotide primers listed in Table S3. Mitochondrial targeting sequences were predicted using MitoProt (Claros, 1995) and were omitted from the constructs. Proteins were expressed as full length (FL) and OB-fold-only polypeptides and contained a cleavable N-terminal hexa-histidine (His_6)-tag: TbMP19-FL (aa 18-170), TbMP63-FL (aa 53-587), TbMP81-FL (aa 56-762), TbMP24-OB (aa 126-246), TbMP24-FL (aa 47-246), TbMP42-OB (aa 245-371), TbMP42-FL (aa 22-371), TbMP63-OB (aa 472-587) and TbMP81-OB (aa 626-762). Details of the

expression and purification are described in the Supplemental Information.

RNA chaperone Assay and Protein Surface Area Calculation

RNA chaperone activity assays of the different protein constructs were performed as described in chapter II (Leeder et al., 2016a), using pre-edited CYb-RNA (1nM) and "gDNA" CYb-5 (100nM) as a pre-mRNA-gDNA pair. Experiments were conducted using minimally 6 different protein concentrations (10-500nM), in at least two independent experiments, to yield ≥12 data points. The amount of cleaved RNA was determined by peak integration and was normalized for each experiment. Editing-active 20S editosomes served as a positive control and were prepared and analyzed as in chapter II (Leeder et al., 2016a). Data points were fitted to a sigmoidal function: $y=L_{low}+L_{up}/[1+e^{(c_{1/2}-x/rate)}]$ (L_{low}=lower limit, L_{up}=upper limit). The concentration of half maximal RNA cleavage ($c_{1/2}$) represents a proxy for the RNA chaperone activity. Normalized surface areas were calculated based on the assumption that the proteins have globular shapes. Volume (V) calculations used an average partial specific volume of 0.73cm^3/g from the molecular masses determined by size exclusion chromatography (Erickson, 2009). Surface areas (A) were calculated as: $A=(36\pi V^2)^{1/3}$. All values were normalized to the surface of TbMP81-OB as the smallest protein.

RNA Synthesis and RNA Structure Probing

Pre-edited transcripts of apocytochrome b (CYb) and of ribosomal protein S12 (RPS12) were generated by runoff *in vitro* T7-transcription following standard procedures. (^{32}P)-labeled RNA preparations were generated by adding α-[^{32}P]-UTP to the transcription mix. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) was conducted as in Leeder et al., 2016a using 1-methyl-7-nitroisatoic anhydride (1M7) as the modification reagent. After the RNA refolding step a 150-fold molar excess (300pmol) of TbMP18-FL was added to the reaction mix and incubated at 27°C for 30min. Raw electrophoretic traces were analyzed using SHAPEfinder (Vasa et al., 2008), utilizing the boxplot approach to determine the number of statistical outliers. Normalized SHAPE-reactivities were generated by averaging a minimum of 3 independent experiments and were used as pseudo-Gibbs free energies to calculate RNA 2D-structures for a temperature of 37°C, using RNAstructure v5.6 (Deigan et al., 2009; Reuter et al., 2010). ShapeKnots (Hajdin et al., 2013) was used to search for pseudoknots, using the default parameters p1=0.35 kcal/mol and p2=0.65 kcal/mol.

Structure Modeling of the 20S Editosome

Models of the *T. brucei* 20S editosome structure were generated using PyRy3D (http://genesilico.pi/pyry3d/) using the cryo-EM density map (EMDB:1595) of Golas et al., 2009. Details of the modeling procedure are specified in the Supplemental Information.

References

Aphasizhev, R., and Aphasizheva, I. (2014). Mitochondrial RNA editing in trypanosomes: small RNAs in control. Biochimie 100, 125-131.

Aphasizheva, I., Maslov, D., Wang, X., Huang, L., and Aphasizhev, R. (2011). Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. Mol. Cell 42, 106-117.

Bae, W., Xia, B., Inouye, M., and Severinov, K. (2000). *Escherichia coli* CspA-family RNA chaperones are transcription antiterminators. Proc. Natl. Acad. Sci. USA. 97, 7784-7789.

Bernacchi, S., Stoylov, S., Piémont, E., Ficheux, D., Roques, B.P., Darlix, J.L., and Mély, Y. (2002). HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. J. Mol. Biol. 317, 385-399.

Böhm, C., Katari, V.S., Brecht, M., and Göringer, H.U. (2012). *Trypanosoma brucei* 20S editosomes have one RNA substrate-binding site and execute RNA unwinding activity. J. Biol. Chem. 287, 26268-26277.

Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009). The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. Cell 139, 547-559.

Bokinsky, G., Nivón, L.G., Liu, S., Chai, G., Hong, M., Weeks, K.M., and Zhuang, X. (2006). Two distinct binding modes of a protein cofactor with its target RNA. J. Mol. Biol. 361, 771-84.

Brecht, M., Niemann, M., Schlüter, E., Müller, U.F., Stuart, K., and Göringer, H.U. (2005). TbMP42, a protein component of the RNA editing complex in African trypanosomes, has endo-exoribonuclease activity. Mol. Cell 17, 621-30.

Claros, M.G. (1995). MitoProt, a Macintosh application for studying mitochondrial proteins. Comput. Appl. Biosci. 11, 441-447.

Croitoru, V., Semrad, K., Prenninger, S., Rajkowitsch, L., Vejen, M., Laursen, B.S., Sperling-Petersen, H.U., and Isaksson, L.A. (2006). RNA chaperone activity of translation initiation factor IF1. Biochimie 88, 1875-1882.

Cross, G.A. (1975). Identification, purification and properties of clone-specific glycoprotein antigens

constituting the surface coat of *Trypanosoma brucei*. Parasitology 71, 393-417.

Czerwoniec, A., Kasprzak, J.M., Bytner, P., Dobrychłop, M., and Bujnicki, J.M. (2015). Structure and intrinsic disorder of the proteins of the *Trypanosoma brucei* editosome. FEBS Lett. 589, 2603-2610.

Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. (2009). Accurate SHAPE-directed RNA structure determination. Proc. Natl. Acad. Sci. USA 106, 97-102.

Drozdz, M., Palazzo, S.S., Salavati, R., O'Rear, J., Clayton, C., and Stuart, K. (2002). TbMP81 is required for RNA editing in *Trypanosoma brucei*. EMBO J. 22, 1791-1799.

Duval, M., Korepanov, A., Fuchsbauer, O., Fechter, P., Haller, A., Fabbretti, A., Choulier, L., Micura, R., Klaholz, B.P., Romby, P., Springer, M., and Marzi, S. (2013). *Escherichia coli* ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation. PLoS Biol. 11, e1001731.

Erickson, H.P. (2009). Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. Bio. Proced. Online 11, 32-51.

Göringer, H.U. (2012). 'Gestalt,' composition and function of the *Trypanosoma brucei* editosome. Annu Rev. Microbiol. 66, 65-82.

Golas, M.M., Böhm, C., Sander, B., Effenberger, K., Brecht, M., Stark, H., and Göringer, H.U. (2009). Snapshots of the RNA editing machine in trypanosomes captured at different assembly stages *in vivo*. EMBO J. 28, 766-778.

Guo, X., Ernst, N.L., and Stuart, K.D. (2008).The KREPA3 Zinc Finger Motifs and OB-Fold Domain Are Essential for RNA Editing and Survival of *Trypanosoma brucei*. Mol. Cell. Biol. 28, 6939-6953.

Hajdin, C.E., Bellaousov, S., Huggins, W., Leonard, C.W., Mathews, D.H., and Weeks, K.M. (2013). Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. Proc. Natl. Acad. Sci. USA 110, 5498-503.

Huang, C.E., O'Hearn, S.F., and Sollner-Webb, B. (2002). Assembly and Function of the RNA Editing Complex in *Trypanosoma brucei* Requires Band III Protein. Mol. Cell. Biol. 22, 3194-3203.

Kala, S., and Salavati, R. (2010). OB-fold domain of KREPA4 mediates high-affinity interaction with guide RNA and possesses annealing activity. RNA 16, 1951-67.

Katari, V.S., van Esdonk, L., and Göringer, H.U. (2013) Molecular crowding inhibits U-insertion/-deletion RNA-editing *in vitro*: consequences for the *in vivo* reaction. PLoS One 12, e83796.

Kazakov, A.S., Markov, D.I., Gusev, N.B., and Levitsky, D.I. (2009). Thermally induced structural changes of intrinsically disordered small heat shock protein Hsp22. Biophys. Chem. 145, 79-85.

Korneta, I., and Bujnicki, J.M. (2012). Intrinsic disorder in the human spliceosomal proteome. PLoS Comput. Biol. 8, e1002641.

Kozlowski, L.P., and Bujnicki, J.M. (2012). MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinformatics 13, 111.

Law, J.A., O'Hearn, S., and Sollner-Webb, B. (2007). In *Trypanosoma brucei* RNA editing, TbMP18 (Band VII) is critical for editosome integrity and for both insertional and deletional cleavages. Mol. Cell. Biol. 27, 777-787.

Leeder, W.M., Voigt, C., Brecht, M., and Göringer, H.U. (2016a). The RNA chaperone activity of the *Trypanosoma brucei* editosome raises the dynamic of bound pre-mRNAs. Sci. Rep. 6, 19309.

Leeder, W.M., Hummel, N.F., and Göringer, H.U. (2016b). Multiple G-quartet structures in pre-edited mRNAs suggest evolutionary driving force for RNA editing in trypanosomes. Sci. Rep. 6, 29810.

Leeder, W.M., Reuss, A.J., Brecht, M., Kratz, K., Wachtveitl, J., and Göringer HU. (2015). Charge reduction and thermodynamic stabilization of substrate RNAs inhibit RNA editing. PLoS One. 10, e0118940.

Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., and Conti, E. (2008). Structure of the active subunit of the yeast exosome core, Rrp44. Diverse modes of substrate recruitment in the RNase II nuclease family. Mol. Cell 29, 717-728.

Low, J.T., and Weeks, K. M. (2010). SHAPE-directed RNA secondary structure prediction. Methods 52, 150-158.

McDermott, S.M., Luo, J., Carnes, J., Ranish, J.A., and Stuart, K. (2016). The architecture of *Trypanosoma brucei* editosomes. Proc. Natl. Acad. Sci. USA pii: 201610177.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics 16, 404-405.

Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K. M. (2005). RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl

acylation and primer extension (SHAPE). J. Am. Chem. Soc. 127, 4223-4231.

Moll, I., Leitsch, D., Steinhauser, T., and Bläsi, U. (2003). RNA chaperone activity of the Sm-like Hfq protein. EMBO Rep. 4, 284-289.

Murzin, A.G. (1993) OB (oligonucleotide/oligo-saccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J. 12, 861-867.

Nakaminami, K., Karlson, D.T., and Imai, R. (2006). Functional conservation of cold shock domains in bacteria and higher plants. Proc. Natl. Acad. Sci. USA 103, 10122-10127.

Niemann, M., Brecht, M., Schlüter, E., Weitzel, K., Zacharias, M., and Göringer, H.U. (2008). TbMP42 is a structure-sensitive ribonuclease that likely follows a metal ion catalysis mechanism. Nucleic Acids Res. 36, 4465-4473.

Niesen, F.H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat. Protoc. 2, 2212-2221.

Park, Y.J., Pardon, E., Wu, M., Steyaert, J., and Hol, W.G. (2012). Crystal structure of a heterodimer of editosome interaction proteins in complex with two copies of a cross-reacting nanobody. Nucleic Acids Res. 40, 1828-1840.

Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics 11, 129.

Salavati, R., Ernst, N.L., O'Rear, J., Gilliam, T., Tarun, S. Jr., and Stuart, K. (2006). KREPA4, an RNA binding protein essential for editosome integrity and survival of *Trypanosoma brucei*. RNA 12, 819-831.

Schnaufer, A., Wu, M., Park, Y.J., Nakai, T., Deng, J., Proff, R., Hol, W.G., and Stuart, K.D. (2010). A Protein-protein interaction map of trypanosome ~20S editosomes. J. Biol. Chem. 285, 5282-5295.

Semrad, K., Green, R., and Schroeder, R. (2004) RNA chaperone activity of large ribosomal subunit proteins from *Escherichia coli*. RNA. 10,1855-1860.

Semrad, K. (2011). Proteins with RNA chaperone activity: A world of diverse proteins with a common task-impediment of RNA misfolding. Biochem. Res. Int. 2011, 532908.

Shoemaker, B.A., Portman, J.J., and Wolynes, P.G. (2000). Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc. Natl. Acad. Sci. USA 97, 8868-8873.

Tarun, S.Z. Jr., Schnaufer, A., Ernst, N.L., Proff, R., Deng, J., Hol, W., and Stuart, K. (2008). KREPA6 is an RNA-binding protein essential for editosome integrity and survival of *Trypanosoma brucei*. RNA 14, 347-358.

Tompa, P., and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB J. 18, 1169-1175.

Varadi, M., Zsolyomi, F., Guharoy, M., and Tompa, P. (2015). Functional advantages of conserved intrinsic disorder in RNA-binding proteins. PLoS One 10, e0139731.

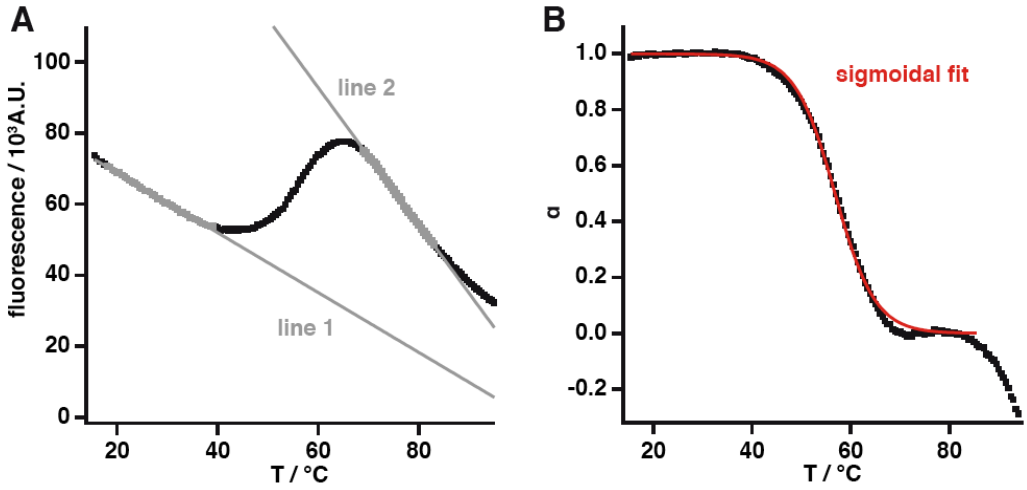
Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M., and Giddings, M.C. (2008). ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA 14, 1979-90.

Worthey, E.A., Schnaufer, A., Mian, I.S., Stuart, K., and Salavati, R. (2003). Comparative analysis of editosome proteins in trypanosomatids. Nucleic Acids Res. 31, 6392-408.

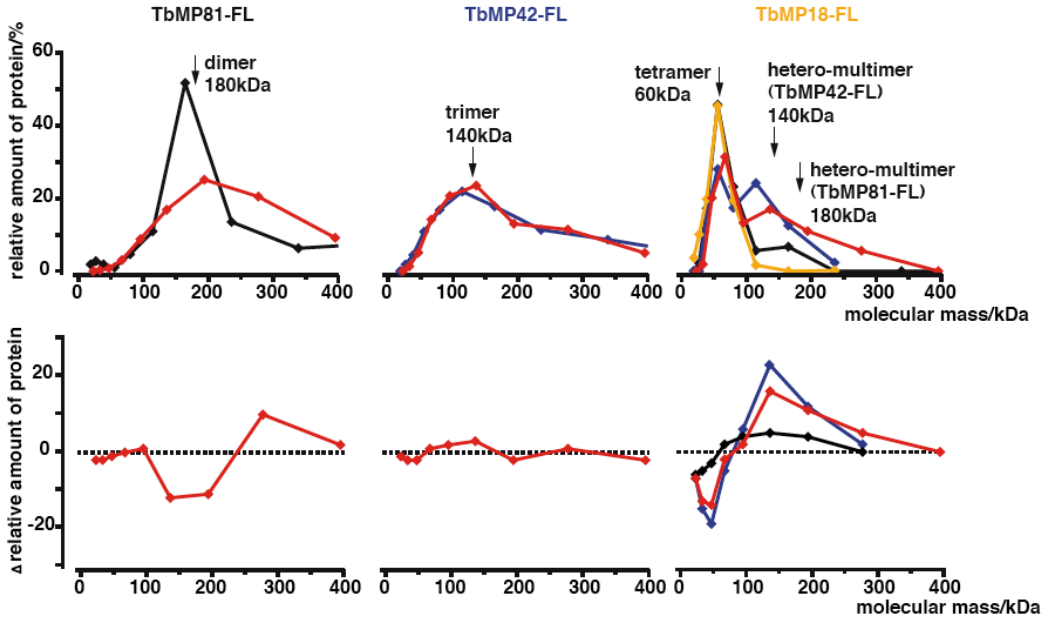
Wu, Z., Hu, G., Yang, J., Peng, Z., Uversky, V.N., and Kurgan, L. (2015). In various protein complexes, disordered protomers have large per-residue surface areas and area of protein-, DNA-, RNA-binding interfaces. FEBS Lett. 589, 2561-2569.

Zhang, A., Derbyshire, V., Salvo, J.L., and Belfort, M. (1995). *Escherichia coli* protein StpA stimulates self-splicing by promoting RNA assembly *in vitro*. RNA 1, 783-793.

Supplemental Information



Supplemental Figure S1 (related to Figure III.2C). Thermal Denaturation Analysis of Editosomal OB-Fold Proteins.
A Representative thermal denaturation profile of the TbMP24-FL-TbMP18-FL complex, monitored by differential scanning fluorimetry (DSF), using the merocyanine dye SYPRO® Orange. Data analysis followed a two-state folding/unfolding model to convert the denaturation curves into fraction-folded (α)-plots (Marky and Breslauer 1987). Baselines for the folded (BL_F; line 1) and the unfolded states (BL_{UF}; line 2) were generated by linear regression and α-values were calculated as: $\alpha=BL_{UF}-f(T)/(BL_{UF}-BL_F)$.
B. Resulting data points were fitted to a sigmoidal function (red line) to derive half-maximal melting transitions (T_m).



Supplemental Figure S2 (related to Figure III.3). Formation of the Ternary TbMP81-FL-TbMP42-FL-TbMP18-FL Complex.
(Upper panel) Molecular mass distributions of the three proteins, after mixing with TbMP81-FL (black), TbMP42-FL (blue), TbMP18-FL (yellow) or a mixture of all three proteins (red). The distributions were generated by quantifying the intensity of protein bands in SDS-containing polyacrylamide gels and plotting the relative quantities as a function of the mean molecular mass. Mean molecular masses were derived from the mean elution volumes of the corresponding fractions.
(Lower panel) Difference plots (hetero-oligomeric distribution minus homo-oligomeric distribution). Coloring scheme as above.

Supplemental Table S1 (related to Figure III.1B).Structure Modeling of the *T. brucei* Editosomal OB-fold Domains.
The published structures of TbMP18, TbMP42 and TbMP81 (Wu et al., 2011; Park et al., 2012; Park and Hol 2012; Park et al., 2012) were used as templates to generate homology models for all OB-fold domains of the *T. brucei* editosome using Swiss Model (Arnold et al. 2006; Kiefer et al. 2009; Guex et al. 2009;Biasini et al., 2014). The modeled protein, the modeling template, the Qmean-score and the sequence of the modeled protein domain are listed.

no.	protein	template pdb	chain	template protein	Qmean	sequence
1	TbMP81	4DKA	C	TbMP81	0.50	SNALMIGRIADVQHGFLGAMTVTQYVLEVDGDERINSKGVTTPASACTPDPASTKAVE AKGEEGEVVE PEKEFIVIRCMGDNFPASLLKDQVKLGSRVLVQGTLRMNRHVDDVSKRLHAYPFIQVVPPLGYVKVV
2	TbMP81	4DNI	A	TbMP18	0.46	DPGSNALMIGRIADVQHGFLGAMTVTQYVLEVDGDERINSKGVTTPASACTPDPASTKAVEAKGEEGE VVEPEKEFIVIRCMGDNFPASLLKDQVKLGSRVLVQGTLRMNRHVDDVSKRLHAYPFIQVVPPLGYVKVV
3	TbMP81	4DK6	C	TbMP81	0.47	SNALMIGRIADVQHGFLGAMTVTQYVLEVDGDERINSKGVTTPASACTPDPASTKAVEAKGEEGEVVE PEKEFIVIRCMGDNFPASLLKDQVKLGSRVLVQGTLRMNRHVDDVSKRLHAYPFIQVVPPLGYVKVV
4	TbMP81	4DK3	C	TbMP81	0.44	SNALMIGRIADVQHGFLGAMTVTQYVLEVDGDERINSKGVTTPASACTPDPASTKAVEAKGEEGEVVE PEKEFIVIRCMGDNFPASLLKDQVKLGSRVLVQGTLRMNRHVDDVSKRLHAYPFIQVVPPLGYVKVV
5	TbMP63	4DNI	A	TbMP18	0.56	IPHAQCVSNITIVGRVLDVSQASENVSHVTVFVEGERSGEEETLTLCCFGEVSQKIRGTLKRNATIFASG TLRLHPVYEASNNKYYVSPVVHVSMPTGTLAVI
6	TbMP42	3STB	C	TbMP42	0.81	AAHWRCVNHCVMLGVVQNIQEGFVFEDKVLQFTLITDFEGPSPGDPDKDFHTVRVFDSDYSSRVKEQ LRDGEWFLVTGRLRMVPQYDGSMRKYYHYPVIQVHPGCGSVLKV
7	TbMP42	4DNI	A	TbMP42	0.71	TILPQAPQYHLDVAPNAPEEGEVAAHWRCVNHCVMLGVVQNIQEGFVFEDKVLQFTLITDFEGPSPGD PDKDFHTVRVFDSDYSSRVKEQLRDGEWFLVTGRLRMVPQYDGSMRKYYHYPVIQVHPGCGSVLKV
8	TbMP24	4DNI	A	TbMP42	0.47	HIEGVVRQVECGYVGGDRVLQFIIEVEEVPASGGTPMRLPLAVRWRPGSEVAQNRMDEWQKEMERL VGRRVLASGRLQVEECFDSGSRRLYKTPSLVLPATSTVE
9	TbMP19	4DNI	A	TbMP18	0.68	TMVGAMHDIQVGFLDRCSVFQFTLTCTVLDFQKVAEPQPKSPGSLPSSTRTAPNANGEEVEKHINKEQ YTVRCLGSEAYTEALKNYLDDGCIVRVIGRLKTTEVVDAGKKQPFPCIIVEQGRWSVSLVHSL
10	TbMP18	4DNI	A	TbMP18	0.82	KSVNSVTLVGVVHDIQSGFVYEDAVTQFTLTTTSIDTTHPTQEVVVEKDHHTIRCFGELFSAEVKQKVK EGNVVCVNGRLRLSPQLEPSCNKHFYFPYIQVQPPHGQVAVIHG
11	TbMP18	3K81	D	TbMP18	0.62	SVNSVTLVGVVHDIQSGFVYEDAVTQFTLTTTSIDTTHPTQEVVVEKDHHTIRCFGELFSAEVKQKVKE GNVVCVNGRLRLSPQLEPSCNKHFYFPYIQVQPPHGQVAVIH
12	TbMP18	3STB	D	TbMP18	0.77	KSVNSVTLVGVVHDIQSGFVYEDAVTQFTLTTTSIDTTHPTQEVVVEKDHHTIRCFGELFSAEVKQKVK EGNVVCVNGRLRLSPQLEPSCNKHFYFPYIQVQPPHGQVAVIHG

Supplemental Table S2 (related to Figure III.1C). Secondary Structure Characteristics of the Consensus OB-Fold Domain.
A consensus structure of the modeled OB-fold elements was generated using MAPSCI (Ilinkin and Janardan, 2010). The protein models are numbered as in Table S1 (numbers 1 to 12). α=α-helix. β=β-sheet and L=connecting loop sequence. Numbers represent the number of amino acids in the different secondary structure elements.

no.	10	11	12	9	8	7	6	5	4	2	3	1	average	sd
β 1	15	15	15	13	13	15	15	15	17	15	17	17	15.2	1.3
L12	2	2	2	2	2	2	2	2	2	2	2	2	2.0	0.0
β 2	14	12	17	16	12	12	12	12	11	12	10	10	12.5	2.2
L23	8	12	3	28	8	8	8	1	40	38	42	42	19.8	16.6
β 3	12	10	15	14	7	10	10	10	8	10	7	7	10.0	2.6
L3α	1	1	0	5	-	1	1	2	-	-	5	5	2.3	2.1
α	10	10	10	7	-	10	10	8	-	-	7	7	8.8	1.5
Lα4	4	5	4	4	-	4	4	4	-	-	5	5	4.3	0.5
β 4	16	11	12	16	14	14	16	16	16	16	14	14	14.6	1.7
L45	4	12	12	2	8	8	5	4	4	4	8	8	6.6	3.3
β 5	10	6	6	10	7	8	10	10	10	10	8	8	8.6	1.6
L56	5	5	5	6	5	5	5	5	5	5	5	5	5.1	0.3
β 6	5	5	5	5	3	5	5	5	5	5	5	5	4.8	0.6
not analyzed	-	-	-	-	25	-	-	-	17	16	-	-	-	-
sum β	72	59	70	74	56	64	68	68	67	68	61	61	65.7	5.5
sum L	24	37	26	47	48	28	25	18	51	49	67	67	40.6	16.7
sum α	10	10	10	7	0	10	10	8	0	0	7	7	8.8	1.5
sum all	106	106	106	128	104	102	103	94	135	133	135	135	115.6	16.0

Supplemental Table S3 (related to Experimental Procedures). DNA-Primer Sequences for the PCR-Amplification and Cloning of the Six Editosomal OB-Fold Proteins.

TbMP81-FL_for	ACTAAGATCTATCAACAACGGACAAGAAGTTTCA
TbMP81-OB_for	CGAGGATCCAGGGAGTAATGCGCTAATG
TbMP81_rev	GATGCGGCCGCTTACCCTACAACTTTCACATAG
TbMP63-FL_for	AGAGGATCCAACGATGGTCGGGGCTATG
TbMP63-OB_for	CGAGGATCCATCTGCCATCGGAGCTTCA
TbMP63_rev	AGTAAGCTTCATTGTGATCACAGCTAGAGTCC
TbMP42-FL_for	ATAGGATCCATCGACTTACGCATCAC
TbMP42-OB_for	ATAGGATCCAACTGGCAGATCAGC
TbMP42_rev	ATACTCGAGTCACACCTTCAACACTG
TbMP24-FL_for	ATAGGATCCAGTCGATGTGGCATATG
TbMP24-OB_for	GAGGATCCAACTGTGACGTCA
TbMP24_rev	AGACTCGAGTTAACACTCCAACTCCTG
TbMP19_for	AGAGGATCCAACGATGGTCGGGGCTATG
TbMP19_rev	TCTCTCGAGTCACTCCAACGTAGCGAC
TbMP18_for	TATCCATGGAGAAGAGTGTAAACAGTGTGAC
TbMP18_rev	TATCTCGAGTTACGATGGCACACCAC

Supplemental Table S4 (related to Figure III.2A). Protein Yield, Solubility and Purity of all Recombinantly Expressed OB-Fold Protein Constructs.

protein	solubility	yield / 1L culture	purity
TbMP18-FL	100%	20mg	97%
TbMP19-FL	0%	1mg	75%
TbMP24-FL/TbMP18-FL	5%	5mg	91%
TbMP24-OB	100%	40mg	94%
TbMP42-FL	30%	6mg	77%
TbMP42-OB	50%	30mg	95%
TbMP63-FL	<1%	0.5mg	53%
TbMP63-OB/TbMP18-FL	30%	5mg	84%
TbMP81-FL	80%	5mg	72%$_1$
TbMP81-OB	15%	8mg	79%

Supplemental Table S5 (related to Figure III.6). Distance Restraints Used in the Structure Modeling of the *T. brucei* 20S Editosome.

Fragment 1	Fragment 2	Maximal Distance
Any Cα atom in fragment 22-100 of TbMP18 (OB-fold)	Any Cα atom in fragment 104-218 of TbMP24 (OB-fold)	15 Å
Any Cα atom in fragment 22-100 of TbMP18 (OB-fold)	Any Cα atom in fragment 291-393 of TbMP42 (OB-fold)	15 Å
Any Cα atom in fragment 22-100 of TbMP18 (OB-fold)	Any Cα atom in fragment 492-588 of TbMP63 (OB-fold)	15 Å
Any Cα atom in fragment 492-588 of TbMP63 (OB-fold)	Any Cα atom in fragment 291-393 of TbMP42 (OB-fold)	15 Å
Any Cα atom in fragment 1-491 of TbMP63 (N-terminal disorder)	Any Cα atom of TbMP52	15 Å
Any Cα atom in fragment 1-491 of TbMP63 (N-terminal disorder)	Any Cα atom of TbMP99 N-terminal domain	15 Å
Any Cα atom in fragment 1-491 of TbMP63 (N-terminal disorder)	Any Cα atom of TbMP99 C-terminal domain	15 Å
Any Cα atom of TbMP44	Any Cα atom in fragment 291-393 of TbMP42 (OB-fold)	15 Å
Any Cα atom in TbMP19	Any Cα atom in TbMP18	30 Å
Any Cα atom in TbMP19	Any Cα atom in fragment 492-588 of TbMP63 (OB-fold)	30 Å
Any Cα atom in TbMP19	Any Cα atom in exonuclease domain of TbMP100	30 Å
Any Cα atom in TbMP18	Any Cα atom of the exonuclease domain in TbMP99	30 Å
Any Cα atom in TbMP18	Any Cα atom in fragment 1-217 of TbMP90 (N-terminal disorder)	30 Å
Any Cα atom in TbMP18	Any Cα atom in exonuclease domain of TbMP100	30 Å
Any Cα atom in fragment 291-393 of TbMP42 (OB-fold)	Any Cα atom in fragment 1-217 of TbMP90 (N-terminal disorder)	30 Å
Cα atom of TbMP99 N-terminal domain's residue 600	Cα atom of TbMP99 C-terminal domain's residue 1	3.5 Å

Supplemental Experimental Procedures

Expression of Recombinant Proteins

DNA sequences for all editosomal OB-fold protein constructs were PCR-amplified from *T. brucei* genomic DNA (strain Lister427) (Cross 1975), using the DNA oligonucleotide primers listed in Table S3. Plasmid expression vectors were modifications of pCDF-1b and pET-33b (Novagen) with the original protease cleavage sites replaced by the cleavage site for tobacco etch virus (TEV) protease. Mitochondrial targeting sequences were predicted using MitoProt (Claros, 1995) and were omitted from the constructs. The proteins were expressed as full length (FL) and OB-fold-only polypeptides and contained a cleavable N-terminal His_6-tag: TbMP19-FL (aa 18-170), TbMP63-FL (aa 53-587), TbMP81-FL (aa 56-762), TbMP24-OB (aa 126-246), TbMP24-FL (aa 47-246), TbMP42-OB (aa 245-371), TbMP42-FL (aa 22-371), TbMP63-OB (aa 472-587) and TbMP81-OB (aa 626-762). The pRSF-duet plasmid for the expression of His_6-tagged TbMP18-FL (aa 19-164) was a gift from Wim Hol (University of Washington, Seattle). Recombinant plasmids were propagated in *E. coli* DH5α. For protein expression, the plasmids were transformed into Rosetta™(DE3)pLysS (Novagen). Cells were grown to late log phase (A_{600} app. 0.8) in Luria broth at 37°C. Protein expression was induced by adding 1mM isopropyl 1-thio-β-D-galactopyranosid (IPTG) at 18°C overnight or at 37°C for 4h. Cells were harvested by centrifugation and processed either directly or stored at -20°C until further use.

Protein Purification and Size Exclusion Chromatography

E. coli cells were lysed in the presence of 0.25mg/mL lysozyme by 3 freeze/thaw cycles and sonication in binding buffer (BB): 20mM Tris/HCl, pH 7.6, 0.3M NaCl, 20mM imidazole containing 1mM phenylmethanesulfonyl fluoride (PMSF) and 1µg/mL leupeptin. In the case of the Zn-finger-containing proteins (TbMP42, TbMP63, TbMP81) 0.1mM $ZnSO_4$ was added. Proteins were purified by immobilized metal affinity chromatography (IMAC) using a Ni-Sepharose HP-resin (GE Healthcare). Protein binding was performed in batch. All subsequent washing and elution steps (with 250mM imidazole in BB) were performed "in-column". His_6-tags were removed by treating the eluates with TEV-protease and 1mM DTT overnight (4°C) followed by a second binding step to the affinity-resin. Final protein solutions were supplemented with DTT (2mM), shock frozen in liquid nitrogen, either directly or after adding 50% (v/v) glycerol and stored at -80°C. The oligomerization state of the different protein constructs was analyzed by size exclusion chromatography using a Superdex 200 HR 10/30 column in 20mM Tris/HCl pH 7.6, 300mM NaCl, 1mM DTT. Elution profiles were monitored by absorbance measurements at 280nm (A_{280}). Molecular masses were determined by calibration to globular proteins. Protein fractions were precipitated with 10% (w/v) trichloroacetic acid (TCA) and analyzed in 10-15% (w/v) SDS-containing polyacrylamide gels.

OB-Fold Structure Prediction

The published structures of TbMP18, TbMP42 and TbMP81 (Wu et al., 2011; Park et al., 2012a; Park and Hol 2012; Park et al., 2012b) were used as templates to generate homology models for the OB-folds of the editosome using Swiss Model (Arnold et al. 2006; Kiefer et al. 2009; Guex et al. 2009; Biasini et al., 2014). A consensus structure of the modeled structures was generated using MAPSCI (Ilinkin and Janardan 2010) and 2D-structures were calculated using PSIPRED (McGuffin et al. 2000). Intrinsic protein disorder propensities were predicted using MetaDisorderMD2 (Kozlowski and Bujnicki 2012).

In vitro Protein Complex Formation

Complexes of editosomal OB-fold proteins were formed *in vitro* in 20mM Tris/HCl pH 7.6, 300mM NaCl and 1mM DTT. Proteins were mixed in 1:1 molar ratios or in the case of TbMP18-TbMP81-FL and TbMP18-TbMP42-FL in a 2:1 molar ratio. Incubation was at 20°C (20min). The formed complexes were analyzed by size exclusion chromatography in combination with gel electrophoresis in SDS-containing polyacrylamide gels. Hetero-oligomeric peak fractions were quantified to determine the approximate ratio of interacting proteins.

Differential Scanning Fluorimetry

The thermal stability of the purified proteins was analyzed by differential scanning fluorimetry (DSF) (Pantoliano et al. 2001, Lo et al. 2004). Experiments were conducted in 10mM Na-cacodylate pH 7.4, 20mM KCl, 10mM $MgCl_2$ and 2mM DTT at a protein concentration of 2µM. For TbMP42-OB a concentration of 17µM was used. Proteins were equilibrated at 15°C for 3min and heated to 95°C with a heating rate of 0.6°C/min. Denaturation was monitored using the merocyanine dye SYPRO® Orange(λ_{ex} 480nm; λ_{em} 605nm). Data analysis was performed assuming a reversible, two-state unfolding reaction to convert the thermal denaturation curves to fraction-folded (α)-plots (Figure S1) (Marky and Breslauer 1987). Baselines for the folded (BL_F) and the unfolded states (BL_{UF}) were generated by linear regression and α-values were calculated as: $\alpha=BL_{UF}–f(T)/(BL_{UF}–BL_F)$. The resulting data points were fitted to a sigmoidal function to derive half-maximal melting points (T_m).

Circular Dichroism Spectroscopy

CD-spectra were recorded between 320nm and 190nm at 27°C in 17mM $K_xH_yPO_4$ pH 7.5, 10mM $MgCl_2$ and 0.5mM DTT at protein concentrations of 0.1-0.2mg/mL in a 0.1cm cuvette. Data were collected every 0.2nm at a scan speed of 50nm/min. Five

spectra were binned and background subtracted. Signals from 190nm-240nm were used to calculate secondary structure contents using CDSSTR (Johnson, 1999) with reference data sets 4 and 7 (Sreerama and Woody, 2000) on the DichroWeb server (Whitmore and Wallace 2004; Whitmore and Wallace 2007). The calculated values of the two reference sets were averaged. The fractions of regular and distorted types of α-helices and β-sheets and the fraction of turns and unordered residues were summed up.

Structure Modeling of the 20S Editosome

Models of the *T. brucei* 20S editosome structure were generated using PyRy3D (http://genesilico.pi/pyry3d/) in combination with the cryo-EM density map of Golas et al., 2009 (EMDB:1595). TbMP18 and TbMP19 were included in the set of proteins. To identify the most likely position of the different proteins, the conformational space was sampled using a Monte Carlo Simulated Annealing method (Metropolis and Ulam, 1949) aiming at maximizing the volume of the map while at the same time minimizing the presence of protein chains outside of the map. Furthermore, restraint violations and steric clashes between the individual components were minimized. Protein sequences predicted to be disordered were modeled in a coarse-grained representation and treated as flexible shapes able to change their conformation during the modeling procedure. PyRy3D was used with its default parameters: Simulated Annealing algorithm, starting temperature T_0=10 in dimensionless units, temperature decrease during the simulation according to $T_n=T_0x0.999^n$ (n=number of the simulation step, 300.000 steps, grid size 2Å). The 0.0419 density threshold was used to define the map volume (Golas et al., 2009). The approximate position of TbMP52 was derived from the electron density map of the *Leishmania tarentolae* RNA editing complex (Li et al., 2009). TbMP52 was pre-oriented to the base of the apex and was allowed to move freely. Three hundred individual simulations were performed. The 100 top-scored models were clustered according to their RMSD-values to obtain groups of similar solutions. The quality of the fit of the models to the cryo-EM map was measured using the cross-correlation coefficient implemented in the "Fit In Map"-procedure of the UCSF Chimera viewer (Pettersen et al., 2004). Superposition of electron density maps was performed using ADP_EM (Garzon et al., 2007). Molecular structures were generated using PyMOL_1.5.0.4 (http://www.pymol.org) and Chimera (Pettersen et al., 2004). Manipulations were performed in Swiss-PDB Viewer (Guex and Peitsch, 1997) and PyMOL_1.5.0.4.

Order/Disorder Analysis of the Editosome Model

The editosome model was analyzed for the positioning of disordered and structured amino acid residues inside and at the surface of the cryo-EM map of the editosome. For that, the volume of the map (at a contour level of 0.0419) was partitioned into an inner region of high electron density and an outer region of low electron density. The position of a residue in the map was defined by its Cα-atom. The number of structured and disordered amino acids inside the map with contour levels of 0.0419 and 0.249 (which corresponds to 50% of the volume of the 0.0419 contour level), and outside the map were calculated. "Inner" residues were defined as residues positioned inside the 0.249 value. "Outer" residues are residues inside the 0.0419, but outside the 0.249 boundary. In addition, a correction was made to account for residues that are inside the shell, but point to the inner cavity. By using the "vop scale"-tool implemented in UCSF Chimera, density values at each point of the electron density map were multiplied by -1, to generate an "imprint" of the editosome representing the regions of negative electron density (at a contour level of 0.0419). This "imprint" was segmented using the immersed watershed algorithm (Pintilie et al., 2010). All segments not connected to the largest cavity were deleted to obtain the surface of the cavity's volume. Minimal distances of the Cα-atoms to the surface of the obtained cavity were calculated. Structured and disordered residues in each region: "inner-editosome", "outer-editosome" and "outside the map" were divided in two categories with respect to their distance to the cavity surface. The threshold was set to 1.2nm. The structured and disordered residues that are on the editosome shell and in the editosome core were calculated as: shell = outer residues (above threshold) + outside map (above threshold); core = inner residues (below + above threshold) + outer residues (below threshold) + outside map (below threshold).

Supplemental References

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics 22, 195-201.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., and Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, W252-W258.

Garzon, J.I., Kovacs, J., Abagyan, R., and Chacon, P. (2007). ADP_EM: fast exhaustive multi-resolution docking for high-throughput coverage. Bioinformatics 23, 427-433.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714-2723.

Guex, N., Peitsch, M.C., and Schwede, T. (2009). Automated comparative protein structure modeling

with SWISS-MODEL and Swiss-Pdb Viewer: A historical perspective. Electrophoresis 30, S162-S173.

Ilinkin, I., Ye, J., and Janardan, R. (2010). Multiple structure alignment and consensus identification for proteins. BMC Bioinformatics 11, 71.

Johnson, W.C. (1999). Analyzing protein circular dichroism spectra for accurate secondary structures. Proteins 35, 307-312.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Res. 37, D387-D392.

Li, F., Ge, P., Hui, W.H., Atanasov, I., Rogers, K., Guo, Q., Osato, D., Falick, A.M., Zhou, Z.H., and Simpson, L. (2009). Structure of the core editing complex (L-complex) involved in uridine insertion/deletion RNA editing in trypanosomatid mitochondria. Proc. Natl. Acad. Sci. USA 106, 12306-12310.

Lo, M.C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., and Ellestad, G. (2004). Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem. 332, 153-159.

Marky, L.A., and Breslauer, K.J. (1987). Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. Biopolymers 26, 1601-1620.

Metropolis, N., and Ulam, S. (1949). The Monte Carlo method. J. Am. Stat. Assoc. 44, 335-341.

Pantoliano, M.W., Petrella, E.C., Kwasnoski, J.D., Lobanov, V.S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B.A., Lane, P., and Salemme, F.R. (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. J. Biomol. Screen. 6, 429-440.

Park, Y.J., Budiarto, T., Wu, M., Pardon, E., Steyaert, J., and Hol, W.G. (2012). The structure of the C-terminal domain of the largest editosome interaction protein and its role in promoting RNA binding by RNA-editing ligase L2. Nucleic Acids Res. 40, 6966-6977.

Park, Y.J., and Hol W.G. (2012). Explorations of linked editosome domains leading to the discovery of motifs defining conserved pockets in editosome OB-folds. J. Struct. Biol. 180, 362-373.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera - a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605-1612.

Pintilie, G.D., Zhang, J., Goddard, T.D., Chiu, W., Gossard, D.C. (2010). Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. J. Struct. Biol. 170, 427-438.

Sreerama, N., and Woody, R.W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal. Biochem. 287, 252-260.

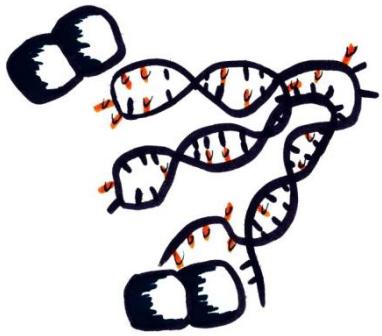
Whitmore, L., and Wallace, B.A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. 32, W668-W673.

Whitmore, L., and Wallace, B.A. (2007). Protein secondary structure analysis from circular dichroism spectroscopy: methods and reference databases. Biopolymers 89, 392-400.

Wu, M., Park, Y.J., Pardon, E., Turley, S., Hayhurst, A., Deng, J., Steyaert, J., and Hol, W.G. (2011). Structures of a key interaction protein from the *Trypanosoma brucei* editosome in complex with single domain antibodies. J. Struct. Biol. 174, 124-136.

Chapter IV

The RNA Chaperone Activity of the Editosomal OB-Fold Protein TbMP24-OB has a Preference for U-Nucleotides



Summary

The editosome of *Trypanosoma brucei* performs an RNA remodeling reaction on mitochondrial transcripts by enhancing the flexibility of primarily U-nucleotides. Since Us are promiscuous base pairing partners, this specificity affects the highest number of possible base pairs. The OB-fold proteins, which are essential for structural integrity of the editosome, contribute to the chaperone surface of the editosome. Here I show that the OB-fold of the protein TbMP24 performs RNA chaperone activity that executes the same nucleotide preference as the complete editosome. It thereby represents a model for the RNA chaperone activity of the fully assembled complex.

Introduction

20S editosomes have been shown to execute an RNA chaperone activity that remodels mitochondrial transcripts of *T. brucei* by increasing the flexibility of primarily U-nucleotides (Us) (Leeder et al. 2016). As Us form base pairs with all other nucleotides, that preference possibly assures that the maximal number of base pairs is affected by RNA remodeling. Although RNA chaperones act in an unspecific manner (Rajkowitsch et al., 2007), similar examples of preference for a distinct nucleotide have been reported before. Two RNA chaperones, hnRNP A1 and a retroviral nucleocapsid protein, were found to act primarily on G-nucleotides (Gs), which are otherwise responsible for the formation of stable intermediate states (Grohman et al., 2013).

The editosomal OB-fold proteins were shown to perform RNA chaperone activities, thereby contributing increments to the RNA chaperone surface of the editosome (chapter III). It remains unclear whether they contribute to the U-preference of the RNA chaperone activity of the editosome or if the detected specificity is contributed by a different editosomal protein.

I investigated the RNA chaperone activity of the OB-fold of the editosomal protein TbMP24 with RNaseH-based gDNA annealing assays and SHAPE-experiments. I found out that the protein has a preference for U-nucleotides and disfavors G-nucleotides. Further, I tested the activity on two different RNAs, the mitochondrial pre-edited RPS12 and CYb transcripts. The detected activity has a preference towards the larger CYb transcript. This is congruent with the substrate preferences of 20S editosomes. The experiments indicate that a single editosomal OB-fold protein performs an RNA chaperone activity with the same substrate preferences as 20S editosomes.

Results

To investigate the specificity of RNA chaperoning, I applied a holistic approach to the transcript of RPS12, which, after pan-editing, encodes the mitochondrial ribosomal protein S12. I synthesized ten gDNAs complementary to RNA regions in different structure and sequence contexts. The gDNAs were designed to obtain gDNA-RNA duplexes of similar thermodynamic stabilities. I titrated the gDNAs in RNaseH-based gDNA annealing assays with RPS12 to determine the concentrations where half-maximal RNaseH cleavage is achieved ($c_{1/2}$) (Figure IV.1B). The values cover a range of three orders of magnitude.

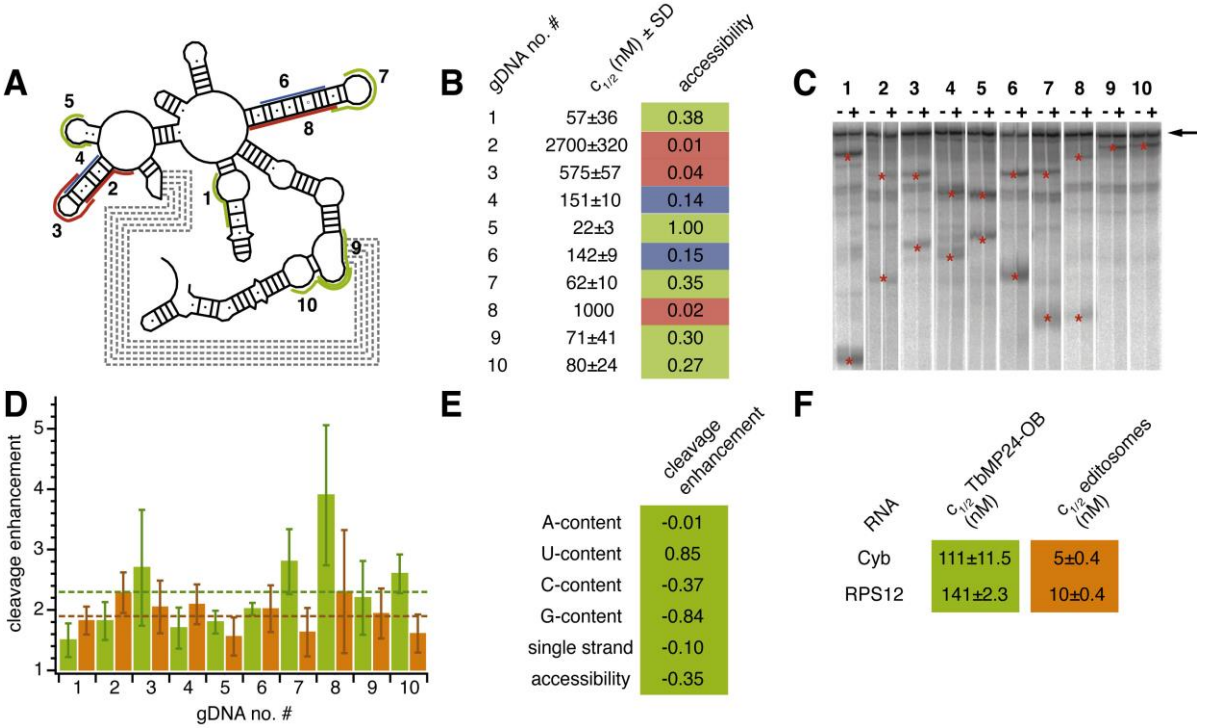


Figure IV.1. Specificity of the RNA Chaperone Activity of TbMP24-OB Probed by RNaseH-Based gDNA Annealing Assays.
A. SHAPE-derived RPS12-RNA structure (Leeder et al., 2016b) with indicated sites covered by gDNAs. Colors indicate accessibility for gDNA binding and RNaseH cleavage of the corresponding RNA region. Green: high accessibility (0.2-1); blue: medium accessibility (0.1-0.2); red: low accessibility (<0.1).
B. gDNA concentrations of half-maximal cleavage ($c_{1/2}$) and derived relative accessibilities of the corresponding RNA regions. Coloring-scheme as in A.
C. Example of autoradiograph of 8M-urea containing polyacrylamide gel of RNaseH cleaved RPS12-RNA with (+) and without (-) TbMP24-OB using gDNAs 1-10. The arrow indicates full length RPS12-RNA and red asterisks indicate the cleavage products.
D. Enhancement of RPS12 cleavage after addition of TbMP24-OB (green) and 20S editosomes (brown) normalized to cleavage without protein (1). The values are averages ± standard deviation. The overall average cleavage enhancement is represented as dashed line.
E. Pearson correlation coefficients of the cleavage enhancement with TbMP24-OB with the number of different nucleotides, relative amount of single stranded nucleotides and accessibility of the corresponding RNA region.
F. RNA-chaperone activity of TbMP24-OB and 20S editosomes probed on two different RNAs.

The $c_{1/2}$ is an inverse measure for the accessibility of the pre-mRNA for annealing and subsequent RNaseH cleavage. I normalized the accessibilities to the maximal value (obtained with gDNA 5) and categorized the resulting normalized accessibilities in three groups representing high, medium and low accessible RNA regions (see Figure IV.1A, B). The accessibilities largely reflect the values expected on basis of the SHAPE-derived secondary structure (Figure IV.1A), except for position three, where a loop shows a low accessibility.

I performed assays utilizing the individual gDNAs to analyze the extent of the RNA chaperone activity of 20S editosomes and TbMP24-OB at the different positions of the RNA. Typical results are shown in Figure IV.1C. RNA chaperone activity of TbMP24-OB and 20S editosomes led to an increase of RNaseH cleavage of the RPS12 transcript with every gDNA (Figure IV.1D). While TbMP24-OB caused a mean cleavage enhancement of 2.3x, with 20S editosomes it was only 1.9x. The variation in cleavage enhancement between the different positions in the RNA is larger with TbMP24-OB than with 20S editosomes. I analyzed the activity of TbMP24-OB for a dependence on the nucleotide composition and the structure of the corresponding RNA region by calculating Pearson correlation coefficients (*r*) (Figure IV.1E). There is no correlation with structural features like the fraction of single-stranded nucleotides or the accessibility, but a correlation with the number of Us in the RNA (*r* 0.85) and an anti-correlation with the number of Gs (*r* -0.84). As the U- and the G-content in the chosen RNA regions is interdependent (*r* -0.92), the dominating determinant cannot be deduced from this experiment.

To check whether there is a preference for a distinct RNA, I performed titrations of TbMP24-OB and 20S editosomes in the RNaseH-based gDNA annealing assay on the RPS12-transcript, in a fashion comparable to experiments with CYb-RNA (chapter II). The determined $c_{1/2}$-values show that TbMP24-OB and 20S editosomes perform RNA chaperone activities on both RNAs with protein concentrations staying in the same order of magnitude (Figure IV.1F). Using the four times larger CYb, 20% less TbMP24-OB and 50% less 20S editosomes are needed to reach the half-maximal effect.

To evaluate the U-specificity of TbMP24-OB, the effects on the RPS12-RNA were analyzed by selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE). SHAPE-reactivity profiles of the mitochondrial transcript RPS12, remodeled by TbMP24-OB, are shown in Figure IV.2A. The mean SHAPE-reactivity is 0.41 SU and thus lies in between free RPS12-transcript and editosome-bound RPS12-transcript (see Figure IV.2B). TbMP24-OB led, like 20S editosomes, to an overall enhancement of the SHAPE-reactivities especially at the termini of the RNA (Figure IV.2B, C, D). The positions with a changed SHAPE-reactivity were analyzed for their nucleotide identity (Figure IV.2E). TbMP24-OB affects thirty percent more U-positions than expected for an unbiased activity. In contrast, it disfavors G-nucleotides, while those that are affected become less flexible. Thus, the SHAPE-experiment confirms the nucleotide-preferences of TbMP24-OB that were monitored by the RNaseH-based gDNA annealing assay.

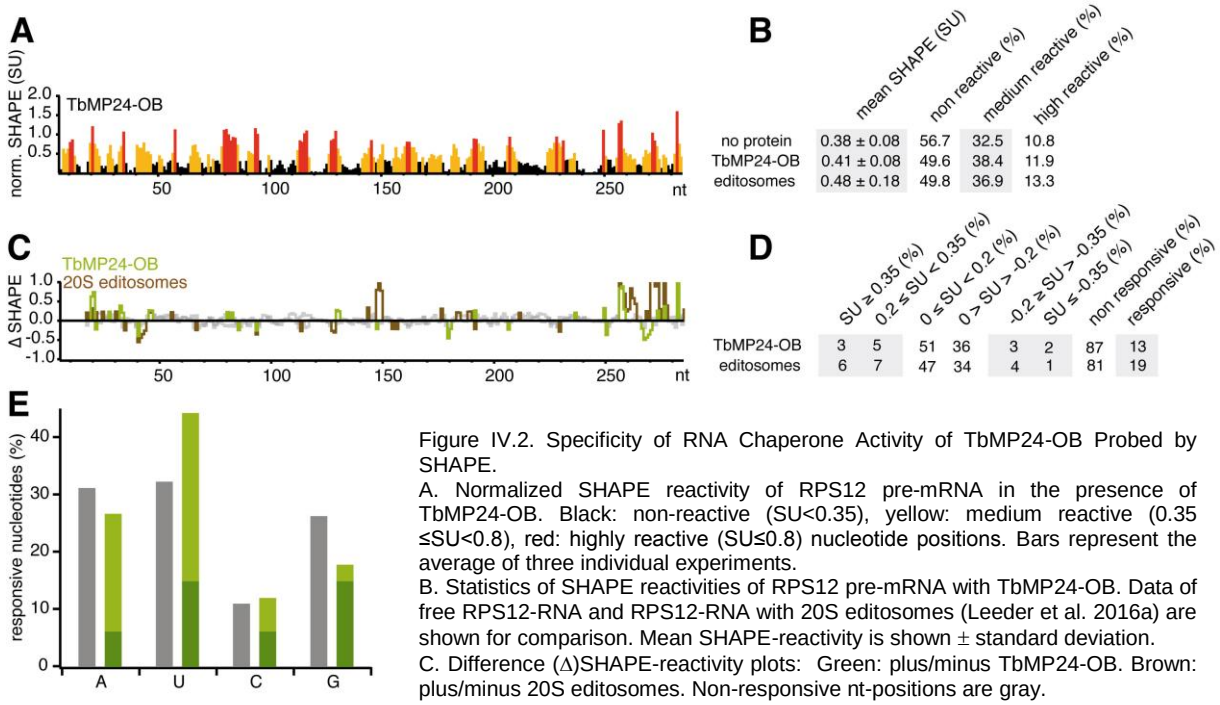


Figure IV.2. Specificity of RNA Chaperone Activity of TbMP24-OB Probed by SHAPE.

A. Normalized SHAPE reactivity of RPS12 pre-mRNA in the presence of TbMP24-OB. Black: non-reactive (SU<0.35), yellow: medium reactive (0.35 ≤SU<0.8), red: highly reactive (SU≤0.8) nucleotide positions. Bars represent the average of three individual experiments.

B. Statistics of SHAPE reactivities of RPS12 pre-mRNA with TbMP24-OB. Data of free RPS12-RNA and RPS12-RNA with 20S editosomes (Leeder et al. 2016a) are shown for comparison. Mean SHAPE-reactivity is shown ± standard deviation.

C. Difference (Δ)SHAPE-reactivity plots: Green: plus/minus TbMP24-OB. Brown: plus/minus 20S editosomes. Non-responsive nt-positions are gray.

D. Statistics of differential SHAPE reactivities.

E. Nucleotides responsive to RPS12 pre-mRNA remodeling by TbMP24-OB. Gray bars: relative nucleotide composition of RPS12, green bars: relative nucleotide composition of the reactive nucleotides, with light green: nucleotides of increased reactivity, dark green: nucleotides of decreased reactivity.

Discussion

To test whether the RNA chaperone activity exhibits substrate preferences, I applied RNaseH-based RNA chaperone assays. Therefore, I utilized gDNAs covering regions of the RPS12-transcript with different sequence and structural features. The gDNAs were titrated to determine the concentration range suitable for the chaperone assay. RNaseH-cleavage can be used as a measure to determine single-stranded regions and thereby accessibility for deoxyoligonucleotide binding (Kauffmann et al., 2009), although the 3D arrangement of the RNA can sterically limit the accessibility for RNaseH cleavage (Matveeva et al., 1997). The accessibilities of the RPS12-RNA regions towards RNaseH-cleavage largely reflect the SHAPE-derived secondary structure of the RNA. Conflicts map to a region where gDNA RPS12-3 binds at the loop of a stem-loop structure. Although half of the nucleotides are single-stranded, the accessibility is in the range of the least accessible regions in double-strands. RNaseH-cleavage in this region is likely sterically hindered by the spatial arrangement of the RNA.

Investigation of TbMP24-OB and the 20S editosome-mediated RNA chaperone activity with the set of gDNAs showed that both protein complexes act qualitatively identical, by increasing the gDNA annealing competence in every region (Figure IV.1D). While the activity of 20S editosomes on the RPS12-RNA results in individual effects fluctuating marginally around the average value, TbMP24-OB leads to pronounced differential effects. Regarding the specifications of the corresponding RNA regions, it is obvious that the differences in TbMP24-OB RNA chaperone activity are due to a preference for Us or a disfavoring of Gs, while the structural context had no effect. The same preferences were confirmed in SHAPE-experiments. A preference to make Us more and Gs less flexible is in line with RNaseH-based gDNA annealing assays. The result points towards similar nucleotide preferences of the editosome protein TbMP24-OB and 20S editosomes, although the preference seems to be more pronounced for TbMP24-OB.

To compare the RNA chaperone activity on different RNA substrates, I titrated TbMP24-OB and 20S editosomes with RPS12- and CYb-RNAs. The mitochondrial transcript of RPS12 is about two hundred nucleotides long and gets pan-edited in *T. brucei* (Read et al., 1992), while the CYb-transcript gets only marginally edited (Feagin et al., 1987) and is four times larger. Both, TbMP24-OB and 20S editosomes, act on the two transcripts with each similar $c_{1/2}$ values, performing slightly better on the larger CYb-transcript. Although further validation is needed, this could be due to the larger target provided for the initial collision, pointing to the initial contact presenting the rate limiting step.

The experiments show that a single component of the multiprotein editosome TbMP24-OB not only shows a similar quantitative and qualitative RNA chaperone activity (chapter III), but also the same nucleotide specificity and action on different RNAs. It resembles a simplified version of a complex providing an RNA chaperone activity similar to that of the editosome.

Experimental Procedures

RNaseH-Based gDNA Annealing Assay

Assays on the transcript RPS12 were performed and analyzed in analogy to experiments with CYb-transcripts (chapter II, chapter III). The gDNAs used for probing of RPS12 were designed to have T_ms between 30°C and 32°C, except for gDNA RPS12-2. To get any cleavage, the oligonucleotide was elongated to yield a T_m of 36°C. The following DNA oligodeoxynucleotides were used for probing of RPS12, with numbers in brackets indicating the binding region in RPS12: RPS12-1 (76-86): CTC-TTTCTCTC, RPS12-2 (114-126): CAAAAGAAGC-TCT, RPS12-3 (123-135): CCTTTTATTCAAA, RPS12-4 (131-140): GCCTCCCTTT, RPS12-5 (143-152): ACTCTCCTCC, RPS12-6 (210-221): CTAAA-ATCTCTT, RPS12-7 (224-237): AAAAACATATC-TTA, RPS12-8 (236-250): TAAAAAAAATATTAA, RPS12-9 (277-288): ATAAATGAACCT, RPS12-10 (283-295): ACCAAACATAAAT. The products were analyzed on 8M urea-containing 10% (w/v) polyacrylamide gels. The $c_{1/2}$-values for the different gDNAs were determined by titration of the gDNA concentration (5-10000nM) and sigmoidal fitting of the resulting curve. The accessibilities of the RPS12-RNA regions complementary to the different gDNAs were calculated as $1/c_{1/2}$-value and normalized to the accessibility calculated for the RPS12-5 complementary region. For determination of the cleavage enhancement with protein 1nM RPS12-RNA and one of the gDNAs at the $c_{1/2}$ was used, either with 180nM TbMP24-OB or 13nM 20S editosomes. For the titrations of protein with RPS12-RNA (1nM) gDNA RPS12-10 (80nM) was used. Titrations of protein with CYb-RNA (1nM) were performed with gDNA CYb-5 (100nM).

RNA-synthesis and Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE)

Pre-edited transcripts of apocytochrome b (CYb) and of ribosomal protein S12 (RPS12) were generated by runoff *in vitro* T7-transcription following standard procedures (chapter II, Leeder et al., 2016). (^{32}P)-labeled RNA preparations were generated by adding α-[^{32}P]-UTP to the transcription mix (chapter II, chapter III).

The SHAPE-experiments and analysis was performed as described in chapter II (Leeder et al. 2016). After the refolding of the RNA, the transcript was incubated with TbMP24-OB in a 150x molar excess (300pmol) over the transcript for 30 minutes at 27°C.

References

Feagin, J.E., Jasmer, D.P., and Stuart, K. (1987). Developmentally regulated addition of nucleotides within apocytochrome b transcripts in *Trypanosoma brucei*. Cell 49, 337-45.

Grohman, J.K., Gorelick, R.J., Lickwar, C.R., Lieb, J.D., Bower, B.D., Znosko, B.M., and Weeks, K.M. (2013). A guanosine-centric mechanism for RNA chaperone function. Science 340, 190-195.

Kauffmann, A.D., Campagna, R.J., Bartels, C.B., and Childs-Disney, J.L. (2009). Improvement of RNA secondary structure prediction using RNase H cleavage and randomized oligonucleotides. Nucleic Acids Res. 37, e121.

Leeder, W.M., Voigt, C., Brecht, M., Göringer, H.U. (2016). The RNA chaperone activity of the *Trypanosoma brucei* editosome raises the dynamic of bound pre-mRNAs. Sci. Rep. 6, 19309.

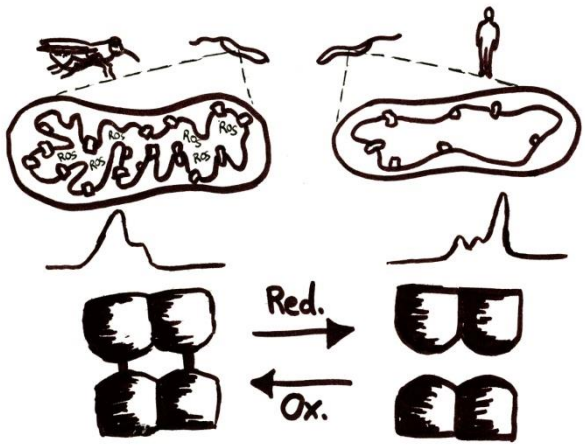
Matveeva, O., Felden, B., Audlin, S., Gesteland, R.F., and Atkins, J.F. (1997). A rapid *in vitro* method for obtaining RNA accessibility patterns for complementary DNA probes: correlation with an intracellular pattern and known RNA structures. Nucleic Acids Res. 25, 5010-5016.

Read, L.K., Myler, P.J., and Stuart, K. (1992). Extensive Editing of Both Processed and Preprocessed Maxicircle CR6 Transcripts in *Trypanosoma brucei*. J. Biol. Chem. 267,1123-1128.

Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M.F., Konrat, R., Bläsi, U., and Schroeder, R. (2007) RNA Chaperones, RNA Annealers and RNA Helicases. RNA Biol. 4, 118-130.

Chapter V

Redox Switching in the *Trypanosoma brucei* Editosomal OB-Fold Proteins



Summary

The RNA editing of transcripts in the mitochondrion of *Trypanosoma brucei* is known to be regulated differentially throughout the lifecycle of the organism. That differential RNA editing affects proteins that are involved in the energy metabolism of the mitochondrion. The cause of differential editing is so far unknown, but there are no differences in the presence of gRNAs or in the protein composition of the editosome. The intracellular location of RNA editing, the mitochondrion, is of differential importance throughout the lifecycle of *T. brucei*. Therefore, it is altered in the presence of respiratory chain proteins, which are the main source of reactive oxygen species, resulting in changes in the redox environment. Here I investigate the response of the integral editosomal OB-fold proteins on an altered redox environment *in vitro*. I show that oxidation leads to cysteine disulfide bridge formation in an editosomal OB-fold protein, which alters the homo- and hetero-oligomerization behavior and thereby possibly the assembly of the editosome.

Introduction

Trypanosoma brucei undergoes a host shift between a tsetse fly and a mammal. Thereby, it adapts to major environmental differences in temperature, nutrient- and oxygen availability by changing morphology, gene expression and energy metabolism. Beside other peculiarities, the organism has just one mitochondrion, which becomes remodeled heavily during the complex life cycle. This remodeling results in a reduced mitochondrial function in *T. brucei* in the mammalian bloodstream. Those parasites generate the required energy solely by glycolysis, whereas procyclic cells in the insect midgut utilize oxidative phosphorylation as the major energy source (Hellemond et al., 2005).

The bloodstream cell mitochondrion lacks most Krebs cycle enzymes, cytochromes and classical respiratory chain proteins. Reducing equivalents generated by glycolysis are transferred to the mitochondrion, where a trypanosomatid alternative oxidase (TAO) serves as electron acceptor. Beside this, a F_0F_1-H^+-ATPase acts under ATP consumption to generate a proton-motive force. The bloodstream cells also contain at least some subunits of a complex I NADH-dehydrogenase. In contrast to that, procyclic cells harbor a fully developed mitochondrion and have a more complex energy metabolism. Although they have all enzymes of the Krebs cycle, only some are used in the conventional way. The respiratory chain is fully developed, comprising respiratory chain complexes I-IV (Hellemond et al., 2005). The functionality of complex I remains elusive and a second type-II NADH dehydrogenase was found, which could serve as an alternative electron entry point to the respiratory chain (Opperdoes and Michels, 2008). Aside from complex IV, procyclic cells also have TAO. The electron transport produces a proton-motive force that is used for ATP synthesis (Hellemond et al., 2005).

As in most eukaryotes, the main source of reactive oxygen species (ROS) in trypanosomatids is the respiratory chain (Tomas and Castro, 2013. Menna-Barreto and de Castro, 2014). Specifically, type-II NADH dehydrogenase was found to be a source of superoxide in procyclic *T. brucei* (Fang and Beattie, 2002). Complex II and III were found to be a source of superoxide in related trypanosomatids (Tomas and Castro, 2013). The absence of most of the respiratory chain proteins in bloodstream trypanosomes results in abolishment of the sources of ROS. This leads to changes in the redox environment in the mitochondrial matrix during the *T. brucei* lifecycle.

The mitochondrion of *T. brucei* is the cellular location of the kinetoplastid specific type of RNA editing, which modifies mitochondrial transcripts by the insertion and/or deletion of exclusively uridylyl residues. That process, is directed by gRNAs and catalyzed by the multiprotein editosome. In some mitochondrial transcripts, RNA editing generates more than half of the nucleotide information, including start and stop codons, whereas other transcripts are not edited at all to generate translatable mRNAs (Göringer, 2012). As the mitochondrial transcripts are encoding proteins that are involved in the energy metabolism, differential RNA editing is a regulation mechanism to generate life-cycle-dependent adaptations of the mitochondrion.

The pre-mRNAs coding for apocytochrome b (complex III) and cytochrome oxidase subunit II (complex IV) are edited preferentially in procyclic cells (Feagin et al., 1987, Feagin et al., 1988), whereas the pre-mRNAs for the NADH dehydrogenase subunits 7 and 8 (complex I) are edited preferentially in bloodstream cells (Koslowsky et al., 1990, Souza et al., 1992). The processes that lead to differential RNA editing are unknown. It was found that the differences are not due to different sets of gRNAs (Koslowsky et al., 1992) and that the editosomes are composed of the same proteins throughout the life cycle (Carnes et al., 2011).

As cysteine thiol redox switching is a mechanism utilized in the regulation of protein function in mitochondria (Nietzel et al., 2016), I tested whether proteins of the *T. brucei* editosome can act as redox switches. I found that three structurally relevant OB-fold proteins of the editosome are sensitive to changes of the redox environment, visible in altered thermal protein stabilities and changes in the oligomerization behavior. I mutated the cysteines of the editosomal OB-fold protein TbMP24. The resulting variant, lacking all cysteines, showed no longer a redox switching behavior, indicative of reduction and oxidation of cysteines being the cause of the switching behavior. The hetero-oligomerization with the interaction partner TbMP18 is not hindered by the cysteine mutations, but the molar ratio of the two interacting proteins changed. These findings have implications for the editosome complex assembly *in vivo*. The data suggest, that cysteine thiol redox switching could be involved in the lifecycle dependent differential editing by influencing the complex architecture under different environmental conditions.

Results

I tested the OB-folds of the editosomal proteins TbMP81, TbMP42 and TbMP24 for their sensitivity towards changes in the redox environment. Therefore I performed size exclusion chromatography and differential scanning fluorimetry (Figure V.1). TbMP42-OB shows a shift in the dominant oligomerization state from a trimer at reducing conditions to a hexamer at non-reducing conditions (Figure V.1A). Differential scanning fluorimetry (DSF) showed that TbMP81-OB becomes thermally stabilized by addition of reducing equivalents, resulting in a ΔT_m of about 10°C (Figure V.1B). TbMP24-OB shows a shift in the size exclusion elution profile from a major dimeric peak at reducing conditions to a major tetrameric peak at non-reducing conditions. The effect could be reversed by addition of reducing equivalents. The reduced form is predominantly a dimer (60%), with two minor species visible, a tetramer (25%) and a trimer (15%). The oxidized form is mainly tetrameric with a minor dimeric species and a background of complexes with indistinct elution behavior (Figure V.1C). I performed DSF to monitor the effects on protein stability. The

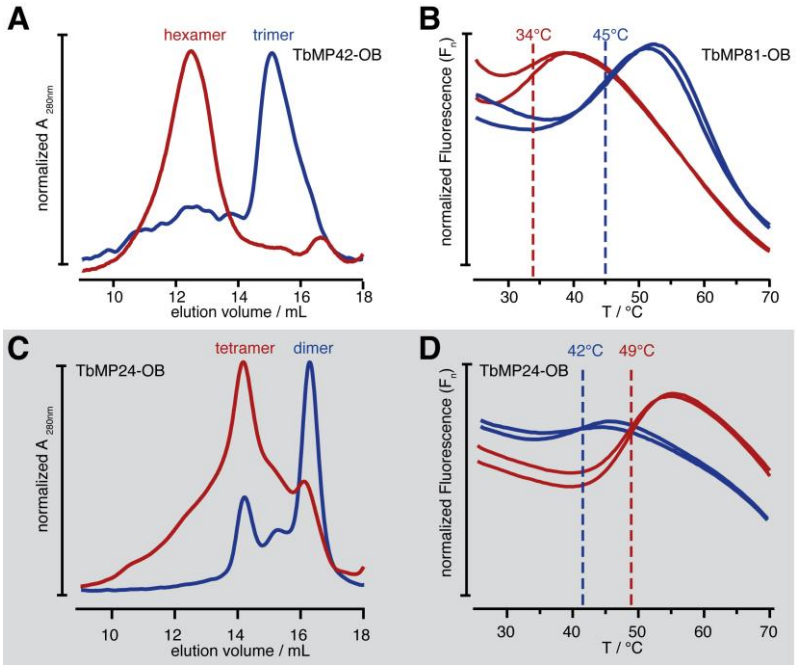


Figure V.1. Redox Sensitivity of *T. brucei* Editosomal OB-Fold Proteins.
A. Gel-filtration chromatography elution profiles of TbMP42-OB under reducing (blue) and non-reducing (red) conditions.
B. Differential scanning fluorimetry of TbMP81-OB under reducing (blue) and non-reducing conditions. The T_ms are indicated as dashed line.
C. Gel-filtration chromatography elution profiles of TbMP24-OB under reducing (blue) and non-reducing (red) conditions.
D. Differential scanning fluorimetry of TbMP24-OB under reducing (blue) and non-reducing (red) conditions. The T_ms are indicated as dashed line.

protein is more stable at non-reducing conditions, with a ΔT_m of 7°C (Figure V.1D). Thus, oxidation of the protein leads to the formation of larger homo-monomeric complexes that have a higher thermal stability.

To test whether disulfide bridge formation between cysteines is responsible for the alterations in the protein stability and oligomerization, I generated mutated proteins lacking two or all four cysteines, respectively. Therefore, I induced mutations with mismatch primers in a PCR and assembled the whole protein by mixing the products and amplifying in a primer-free PCR. In mutant protein 1 the most N-terminal cysteine (C1) was substituted by an alanine and the C-terminal cysteine was removed (C4). The cysteines are in regions at the termini of the protein outside the predicted OB-fold (Figure V.2A). They are at the interface that was found to be involved in tetramerization of TbMP18 and TbMP42 (Park et al., 2012). Mutant protein 2 carries the same mutations of C1 and C4 and additional cysteine to alanine substitutions at position C2 and C3. Cysteine 2 and 3 are predicted to be located at the ends of β-sheets 1 and 4 of the OB-fold, near loops L_{12} and L_{45} (Figure V.2A). These loops are part of the common ligand binding interface of OB-fold proteins (Theobald et al., 2003). I expressed the protein variants recombinantly with an amino-terminal cleavable hexa-histidine tag, as described before for the wild-type (wt) protein (chapter III). The expression and purification resulted in protein preparations of comparable yield, solubility and purity to the wild-type protein (Figure V.2B). The CD-spectra of all three proteins are superimposable and calculation of the secondary structure content resulted in about 60% disorder and about 30% β-sheets for all three proteins (Figure V.2C).

I performed size exclusion chromatography to investigate the oligomerization behavior at reducing and non-reducing conditions (Figure V.2D). Mutant protein 1 shows an identical elution behavior like the wt protein in both conditions, with a distinct redox switching. Regarding the elution profiles of mutant protein 2, they are identical at the different redox conditions. Thus, redox switching is abolished by mutation of C2 and C3. The profiles are similar to those of wt and mutant protein 1 at reducing conditions, but there is none tetramer formation. The results indicate that C2 and/or C3 oxidation leads to redox switching of the oligomerization state of TbMP24-OB from dimer to tetramer.

To investigate whether intermolecular disulfide bridges are also involved in the hetero-oligomer complex formation, I performed *in vitro* interaction experiments of the variants of TbMP24-OB with TbMP18-FL in a 1:1 molar ratio under reducing conditions. Wild-type TbMP24-OB and TbMP18-FL interact by forming a trimeric complex (Figure V.3A, see also Figure III.3). A complex of the same size is found when TbMP18-FL interacts with the two mutated variants. The result of the gel electrophoretic analysis shows, that mutant protein 1 forms the hetero-trimeric complex in the same stoichiometry as wt TbMP24-OB, while in the complex with mutant protein 2 the stoichiometry is altered (Figure V.3B). The distribution of the individual proteins in the fractions is plotted in Figure V.3C. Regarding the elution of TbMP18-FL in the mixing experiments with each TbMP24-OB variant (Figure V.3C, gray curves), there is no difference utilizing the different variants. In each experiment about 10% of the TbMP18-FL protein shifted to a lower molecular mass. Wild-type and mutant 1 TbMP24-OB behave in a similar fashion when interacting with TbMP18-FL. There is approximately a 25% shift from dimeric to tetrameric fractions (Figure V.3C). So the hetero oligomerization in a 2:1 molar ratio of TbMP24-OB and TbMP18-FL is not altered by the mutation of C1 and C4. When all cysteines are mutated, the amount of TbMP24-OB interacting with TbMP18-FL decreased from 25% to 10% of the total protein. So the oligomerization state

of the hetero-oligomeric complex stays the same after removal of all cysteines, but the molar ratio of interacting proteins is altered.

Discussion

The experiments shown that changes in the redox environment can alter the oligomerization state and thermal stability of the editosomal OB-fold proteins. This phenomenon was investigated for TbMP24-OB in detail. The protein reacted to the removal of reducing equivalents by tetramer formation. A small amount of this tetramer was also present at reducing conditions, indicative of an incomplete reduction in the presence of 1mM DTT. The oxidation of TbMP24-OB leads to a thermal stabilization of the homo-oligomer by 7°C. As intra-molecular and inter-molecular disulfide-bonds were found to thermally stabilize protein structures (Trivedi et al., 2009, Nakka et al., 2006), I tested whether the observed effects on the OB-fold protein depend on disulfide bond formation. Mutation of two subsets of cysteines resulted in two protein variants that are – on basis of the observed solubility, the measured CD-spectra and the behavior in a size exclusion chromatography – structurally unaltered compared to wild-type protein. By the analysis of the homo-oligomerization behavior of the mutated proteins, the redox switching could be assigned to one or both of the two internal cysteines. In the structural model, those are in a rigid structural context. The position of the cysteines suggests the site of TbMP24-OB homo-tetramerization. That likely does not happen on the common tetramerization interface (Figure V.2A) that is found in single strand DNA binding proteins (Ollis et al., 1983) and in TbMP18/TbMP42 crystals (Park et al., 2012), but on the common ligand binding interface (Figure V.2A). The oxidation of the cysteines has ramifications for the hetero-oligomerization with the interaction partner TbMP18-FL. The cysteines of TbMP24-OB do not participate directly in the interaction, as hetero-oligomerization is not hindered by the mutations, but change the molar ratio of interacting proteins. So the binding of further monomers is altered by cysteine oxidation, suggesting a competitive effect. A study, which utilized shortened variants of TbMP24, suggested an interaction with TbMP18 via the α-helical region between β-sheets 3 and 4 (Kala et al., 2012). The presented results are in line with that "atypical" oligomerization interface.

Alterations in hetero-oligomerization have implications for the formation of higher ordered complexes. The different redox conditions in the mitochondrion of *T. brucei* throughout its lifecycle could influence the interactions of the editosomal proteins by redox switching. Generally, reactive oxygen species seem to be an important signal for regulation in mitochondria, as many mitochondrial proteins were found to act as redox switches (Mailloux et al., 2013).

Aside from the reported inter-molecular disulfides, intra-molecular disulfides could also have an impact on differential editosome activity. Zn-fingers are known to be sensitive towards oxidation, due to disulfide bridge formation, thereby impairing the Zn^{2+} binding (Larabee et al., 2005, Baldwin and Benz,

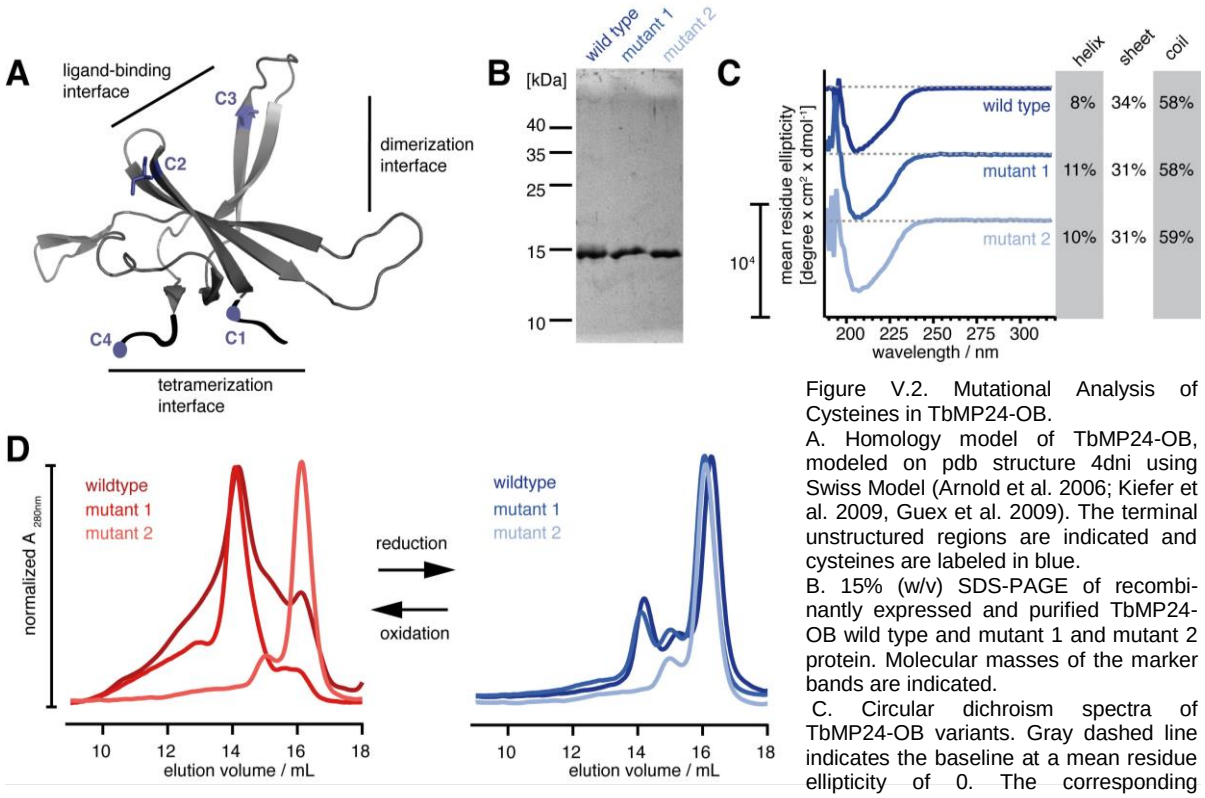


Figure V.2. Mutational Analysis of Cysteines in TbMP24-OB.
A. Homology model of TbMP24-OB, modeled on pdb structure 4dni using Swiss Model (Arnold et al. 2006; Kiefer et al. 2009, Guex et al. 2009). The terminal unstructured regions are indicated and cysteines are labeled in blue.
B. 15% (w/v) SDS-PAGE of recombinantly expressed and purified TbMP24-OB wild type and mutant 1 and mutant 2 protein. Molecular masses of the marker bands are indicated.
C. Circular dichroism spectra of TbMP24-OB variants. Gray dashed line indicates the baseline at a mean residue ellipticity of 0. The corresponding secondary structure contents were calculated using the CDSSTR algorithm (Johnson, 1999).
D. Size exclusion chromatography elution profiles of TbMP24-OB variants under non-reducing (left panel) and reducing (right panel) conditions.

2002). McDermott et al. reported 2015 that mutation of the TbMP42 Zn-fingers and repression of the protein have differential effects on RNA editing and editosome integrity in *T. brucei* bloodstream and procyclic cells. Redox switching, which alters the connectivity of the proteins with the interaction partners, or the Zn-finger structure, would be a plausible explanation for these observations.

Experimental Procedures

Construction of Mutated Protein Variants

The cysteine codons in the gene coding for TbMP24-OB were mutated by site-directed mutagenesis by overlap extension (Ho et al., 1989). As template serves a pCDF plasmid, carrying the gene for wild-type variant TbMP24-OB (chapter III). The mutations were created, using primers that generate two to three mismatches, thereby substituting a cysteine codon by an alanine codon (underlined codons) or deleting the cysteine codon (asterisk indicates position) in a PCR reaction: cys1-ala-fw: GGGGATCCAACTGTGACG-TCAGCCGCCCA, cys2-ala-fw: GCGACAGGTGG-AGGCTGGTTATGTAGGAGG, cys3-ala-fw: TGCA-GGTGGAGGAAGCCTTTGATTCGGGAAG, cys2-ala-rev: CCTCCTACATAACCAGCCTCCACCTGT-CGC, cys3-ala-rev: CTTCCCGAATCAAAGGCTTC-CTCCACCTGCA, cys4-del-rev: TACCAGACTCG-AGTTA*CTCCAACTCCTGCAAAG.
The variant with mutated terminal cysteines was constructed by PCR with primer pair cys1-ala-fw/cys4-del-rev. The template was DpnI-digested, the product purified by phenol extraction and ethanol-precipitated. The variant, with all cysteines mutated, was constructed by PCRs with primer pairs: 1. cys1-ala-fw/cys2-ala-rev, 2. cys2-ala-fw/cys3-ala-rev, 3. cys3-ala-fw/cys4-del-rev. PCR products 1, 2 and 3 were mixed in equimolar amounts and used as template for a second PCR, to assemble the gene containing all mutations. Plasmid DNA was digested with DpnI. A final round of PCR amplification was performed with the primer pair cys1-ala-fw/cys4-del-rev. The final product was gel-purified on a 1% agarose gel and ethanol-precipitated. The mutated genes were cloned into pET-33b and propagated in *E. coli* DH5α. Protein expression in Rosetta™(DE3)pLysS (Novagen) is described in chapter III.

Protein Purification, Size Exclusion Chromatography and CD-Spectroscopy

The protein purification was performed as described in chapter III. Proteins that were analyzed under non-reducing conditions were either dialyzed after the purification against buffer without DTT and stored without DTT or dialysis was performed directly before the experiment (in experiments with TbMP81-OB). Size exclusion chromatography was performed as described in chapter III. Experiments under non-reducing conditions were performed without DTT, experiments under reducing conditions were performed with 1mM fresh DTT in the buffer. The CD-spectroscopy was conducted as described in chapter III.

In vitro Protein Complex Formation

The heteromeric complexes were formed as described in chapter III. The SDS-containing polyacrylamide gels of the eluted fractions were analyzed by quanti-

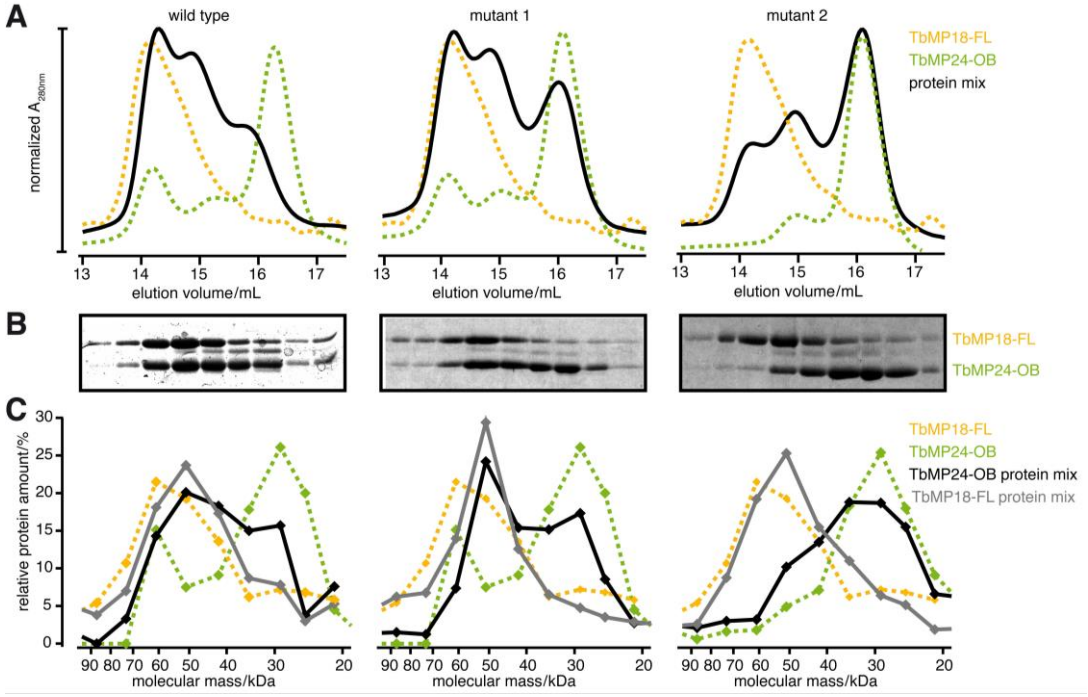


Figure V.3. Interaction of TbMP24-OB Variants with TbMP18-FL.
A. Size exclusion chromatography elution profiles of single proteins and mixing-experiments for each TbMP24-OB variant.
B. 15% (w/v) SDS-containing polyacrylamide gel of 0.5mL fractions corresponding to the mixing experiment above.
C. Relative protein amounts in the 0.5mL fractions of the elution from size exclusion chromatography column, determined from a SDS-containing polyacrylamide gel.

fication of the protein bands. For each protein the total eluted protein in all fractions was normalized to 100%, to generate relative protein molecular mass distributions.

Differential Scanning Fluorimetry (DSF)

DSF was generally performed as described in chapter III, but thermal denaturation curves were monitored starting from 25°C. Experiments were conducted in binding buffer (20mM Tris/HCl, pH 7.6, 0.3M NaCl, 20mM imidazole) +/-2mM DTT with 7µM protein.

References

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics 22, 195-201.

Baldwin, M.A., and Benz, C.C. (2002). Redox control of zinc finger proteins. Methods Enzymol. 353, 54-69.

Carnes, J., Soares, C.Z., Wickham, C., and Stuart, K. (2011). Endonuclease associations with three distinct editosomes in *Trypanosoma brucei*. J. Biol. Chem. 286, 19320-19330.

Fang, J., and Beattie, D.S. (2002). Rotenone-insensitive NADH dehydrogenase is a potential source of superoxide in procyclic *Trypanosoma brucei* mitochondria. Mol. Biochem. Parasitol. 123, 135-142.

Feagin, J. E., Jasmer, D. P., and Stuart, K. (1987). Developmentally regulated addition of nucleotides within apocytochrome b transcripts in *Trypanosoma brucei*. Cell 49, 337-345.

Feagin, J. E., and Stuart, K. (1988). Developmental aspects of uridine addition within mitochondrial transcripts of *Trypanosoma brucei*. Mol. Cell. Biol. 8, 1259-1265.

Göringer, H.U. (2012). 'Gestalt,' composition and function of the *Trypanosoma brucei* editosome. Annu. Rev. Microbiol. 66, 65-82.

Guex, N., Peitsch, M.C., and Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-Pdb Viewer: A historical perspective. Electrophoresis 30, S162-S173.

Hellemond, J.J., Bakker, B.M., and Tielens, A.G. (2005). Energy Metabolism and Its Compartmentation in *Trypanosoma brucei*. Adv. Microb. Physiol. 50, 199-226.

Ho, S.N., Hunt , H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77, 51-59.

Johnson, W.C. (1999). Analyzing protein circular dichroism spectra for accurate secondary structures. Proteins 35, 307-312.

Kala, S., Moshiri, H., Mehta, V., Yip, C.W., and Salavati, R. (2012). The oligonucleotide binding (OB)-fold domain of KREPA4 is essential for stable incorporation into editosomes. PLoS One. 7, e46864.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., and Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. Nucleic Acids Res. 37, D387-D392.

Koslowsky, D. J., Bhat, G. J., Perrollaz, A. L., Feagin, J. E., and Stuart, K. (1990). The MURF3 gene of *T. brucei* contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase. Cell 62, 901-911.

Koslowsky, D. J., Riley, G. R., Feagin, J. E., and Stuart, K. (1992). Guide RNAs for transcripts with developmentally regulated RNA editing are present in both life cycle stages of *Trypanosoma brucei*. Mol. Cell. Biol. 12, 2043-2049.

Larabee, J.L., Hocker, J.R., and Hanas, J.S. (2005). Cys redox reaction and metal binding of a Cys2His2 zinc finger. Arch. Biochem. Biophys. 434, 139-149.

Mailloux, R.J., Jin, X., and Willmore, W.G. (2013). Redox regulation of mitochondrial function with emphasis on cysteine oxidation reactions. Redox Biol. 2, 123-139.

McDermott, S.M., Guo, X., Carnes, J., and Stuart, K. (2015). Differential editosome protein function between life cycle stages of *Trypanosoma brucei*. J. Biol. Chem. 290, 24914-24931.

Menna-Barreto, R.F.S., and de Castro, S.L. (2014). The double-edged sword in pathogenic trypanosomatids: The pivotal role of mitochondria in oxidative stress and bioenergetics. Biomed. Res. Int. 2014, 614014.

Nakka, M., Iyer, R.B., and Bachas, L.G. (2006). Intersubunit disulfide interactions play a critical role in maintaining the thermostability of glucose-6-phosphate dehydrogenase from the hyperthermophilic bacterium *Aquifex aeolicus*. Protein J. 25, 17-21.

Nietzel, T., Mostertz, J., Hochgräfe, F., and Schwarzländer, M. (2016). Redox regulation of mitochondrial proteins and proteomes by cysteine thiol switches. Mitochondrion doi: 10.1016/j.mito.2016.07.010.

Ollis, D., Brick, P., Abdel-Meguid, S.S., Murthy, K., Chase, J.W., and Steitz, T.A. (1983). Crystals of *Escherichia coli* single-strand DNA-binding protein show that the tetramer has D2 symmetry. J. Mol. Biol. 170, 797-800.

Opperdoes, F.R., and Michels, P.A.M. (2008). Complex I of *Trypanosomatidae*: does it exist? Trends Parasitol. 24, 310-317.

Park, Y.J., Pardon, E., Wu, M., Steyaert, J., and Hol, W.G. (2012). Crystal structure of a heterodimer of editosome interaction proteins in complex with two copies of a cross-reacting nanobody. Nucleic Acids Res. 40, 1828-1840.

Souza, A. E., Myler, P. J., and Stuart, K. (1992). Maxicircle CR1 transcripts of *Trypanosoma brucei* are edited and developmentally regulated and encode a putative iron-sulfur protein homologous to an NADH dehydrogenase subunit. Mol. Cell. Biol. 12, 2100-2107.

Theobald DL, Mitton-Fry RM, Wuttke DS. (2003). Nucleic acid recognition by OB-fold proteins. Annu. Rev. Biophys. Biomol. Struct. 32, 115-133.

Tomas, A.M., and Castro, H. (2013). Redox metabolism in mitochondria of trypanosomatids. Antioxid. Redox Signal. 19, 696-707.

Trivedi, M.V., Laurence, J.S., and Siahaan, T.J. (2009).The role of thiols and disulfides in protein chemical and physical stability. Curr. Protein Pept. Sci. 10, 614-625.

Chapter VI
General Discussion

RNA molecules are involved in nearly all biological processes. They execute their function through the interaction with other biomolecular components, foremost with other RNAs and a large number of RNA-interacting proteins. As a consequence, the formation of specific RNA-RNA and RNA-protein complexes is a major task in many biological pathways. For example, the formation of RNA-duplexes between small non-coding (nc)RNAs and their target RNAs is frequently used to define RNA-processing sites. Matchmaker proteins often assist in the search for a binding partner by specifically holding the ncRNA in a binding-competent state (reviewed by Künne et al., 2014). RNA editing in *T. brucei* utilizes such a mechanism: Editing sites are defined as the first non-base-paired nucleotide proximal to a helical gRNA-pre-mRNA duplex. The formation of cognate gRNA-pre-mRNA pairs has been shown to involve proteins with RNA-annealing activity as well as RNA-helicases, which likely catalyze the unwinding of the formed duplex structures to allow for multiple rounds of editing (see chapter I; Kruse at al., 2013). However, the main obstacles for forming productive gRNA-pre-mRNA hybrid RNAs are the highly stable secondary structures of the different pre- and partially edited transcripts (Leeder et al., 2016). In many cases intramolecular duplexes obscure the entry sites for gRNAs and thus represent a rate-limiting barrier for the progression of the reaction (Leeder et al., 2016). In order to resolve these structural barriers, editosomes execute a complex-inherent RNA-chaperone activity. The activity enhances the structural dynamic of the different pre-mRNA molecules (Leeder et al., 2016), thereby generating gRNA-annealing-competent pre-mRNA structures. The molecular nature of the editosome-inherent RNA-chaperone activity was investigated here.

As a first step, I developed an experimental assay to systematically characterize the editosomal RNA chaperone activity. I uncovered that the activity drives the formation of intermolecular duplex structures, thereby promoting the initiation of the RNA editing reaction. The assay was used to test whether the 6 OB-fold proteins of the editosome contribute to the RNA unfolding process. I uncovered that 5 of the proteins are involved in the reaction. I also demonstrate that the RNA remodeling reaction is dependent on the surface areas of the formed complexes, indicating that intrinsically disordered (ID) regions of the different proteins may be involved in the reaction.

Further experiments examining the substrate specificity of the OB-fold protein TbMP24 showed that the polypeptide has the same RNA substrate preferences as the fully assembled editosome. This suggests that the overall binding and chaperoning surface is composed of individual units. Each of which adds a defined activity increment to the overall activity of the assembled complex.

The protein interaction data were used to compute a first coarse-grained model of the editosome. The model suggests that the ID-domains are preferentially located on the surface of the high molecular mass complex, perhaps acting as a multifunctional RNA capturing, binding and remodeling domain (Figure III.7 and VI.1). This model involves intrinsically disordered protein regions as functional units, which on a first glance seems to violate the classic structure-function paradigm. However, spliceosomes have been shown to have a similar organization, with a core of well-structured catalytic proteins and a disordered shell (Korneta and Bujnicki, 2012). In the last years, many protein complexes with such functionally relevant ID protein regions have been identified and termed "fuzzy complexes" (reviewed in Sharma et al., 2015).

The parallels between *T. brucei* editosomes and other RNA-binding proteins and complexes indicate that RNA-binding reactions to intrinsically disordered protein regions might represent a general strategy for the binding and unfolding of substrate RNAs. A disordered surface has several beneficial characteristics. It allows multiple contacts during the binding reaction and enables sufficient conformational flexibility to target different RNA-ligands (Varadi et al., 2015) or perhaps one RNA-ligand with different structures. IDs provide a large "reach" (Shoemaker et al., 2000) to facilitate the binding reaction, thereby loosening the structure, using an entropy transfer mechanism (Tompa and Csermely, 2004). These properties of ID-regions mediate the function of these dynamic "structural" domains in an equally dynamic process.

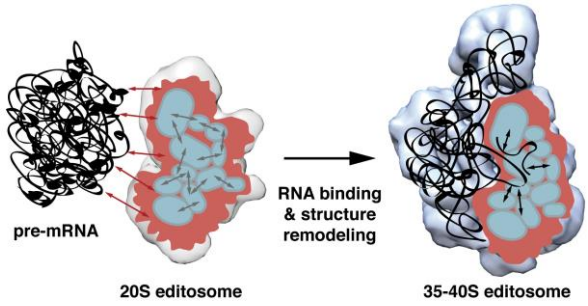


Figure VI.1. Proposed Model of the Interplay Between Order and Disorder During the Editosome-RNA Binding Reaction.
Light blue: structured editosome protein domains; gray arrows: protein-protein interactions; red: intrinsically disordered editosome shell; red arrows: interaction of disordered regions with pre-mRNA (black); black arrows: interactions between pre-mRNA/gRNA and structured protein domains.

In contrast to this model, the 20S editosome was found to bind RNA substrates tightly with low nanomolar dissociation constants (Böhm et al., 2012). This seems to be contradictory to an unspecific RNA binding surface, but it can be explained by the presence of two RNA binding modes. Two binding modes were shown for the interaction of CBP2 with the bI5 group I intron, where an initial rapid and unspecific binding mode induces conformational

fluctuations in the RNA and a slow specific mode stabilizes the native RNA structure. This enhances the chance of two rare events, the formation of a binding-competent RNA structure and the productive collision between the RNA and the protein (Bokinsky et al., 2006). Two binding modes can thereby funnel the binding of the RNA to the protein. Similarly, a continuous funneling was found at distinct RNA-binding sites in a protein. Thereby, nucleic acid binding follows a dynamic process, with an initial binding at suboptimal locations and a free energy funnel driving the RNA to the optimal binding site (Miao and Westhof, 2015).
The probability of pre-mRNA binding to the editosome could be further enhanced by a spatial coupling of RNA processing steps, which creates an RNA processing microenvironment inside the mitochondrion. RNA editing substrates, the editosome and the gRNA binding complex were found to be associated with the large ribosomal subunit (Aphasizheva et al., 2011) and an association to the inner mitochondrial membrane has been discussed by Katari et al. (2013). Microscopic colocalization in closely related *Leishmania tarentolae* showed that editosome components, gRNA binding complex components and ribosomal proteins are concentrated in kDNA-adjacent nodes. This shows that the gRNA-binding complex is closely associated to the two other complexes and all are located near the side of transcription (Wong et al., 2015).

Further, I investigated whether the integral editosomal OB-fold proteins could also serve to switch RNA editing during the lifecycle of the parasite. I found that the assembly of OB-fold complexes can be altered by cysteine oxidation in the structured OB-fold domain. This shows that the arrangement of the OB-fold core structure can be environmentally influenced. Due to the modular assembly of the RNA chaperone surface, an altered connectivity would likely not alter the chaperone activity. The experiment gives a plausible explanation for differential RNA editing by a complex responding to its chemical environment. So, protein complexes with the same overall protein complex composition (Carnes et al., 2011) could perform differential tasks *via* different spatial assemblies of the structural protein domains. This indicates that the editosome is a plastic complex that is influenced by its environment.

The work presented here provides new insights into the architecture of the *T. brucei* editosome and editosome-RNA interactions. The proposed model points towards a highly dynamic RNA binding surface, which acts as RNA chaperone, and a structural core. Experiments indicate that the assembly of the structural core might be regulated differentially through the *T. brucei* lifecycle. This observed organization of structured and intrinsically disordered protein domains could be a general strategy for the organization of RNA-processing protein complexes.

References

Aphasizheva, I., Maslov, D., Wang, X., Huang, L., and Aphasizhev, R. (2011). Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes. Mol. Cell 42, 106-117.

Böhm, C., Katari, V.S., Brecht, M., and Göringer, H.U. (2012). *Trypanosoma brucei* 20S editosomes have one RNA substrate-binding site and execute RNA unwinding activity. J. Biol. Chem. 287, 26268-26277.

Bokinsky, G., Nivón, L.G., Liu, S., Chai, G., Hong, M., Weeks, K.M., and Zhuang, X. (2006). Two distinct binding modes of a protein cofactor with its target RNA. J. Mol. Biol. 361, 771-784.

Carnes, J., Soares, C.Z., Wickham, C., and Stuart, K. (2011). Endonuclease associations with three distinct editosomes in *Trypanosoma brucei*. J. Biol. Chem. 286, 19320-19330.

Katari, V.S., van Esdonk, L., and Göringer, H.U. (2013) Molecular crowding inbitis U-insertion/-deletion RNA-editing *in vitro*: consequences for the *in vivo* reaction. PLoS One 12, e83796.

Korneta, I., and Bujnicki, J.M. (2012). Intrinsic disorder in the human spliceosomal proteome. PloS Comput. Biol. 8, e1002641.

Kruse, E., Voigt, C., Leeder, W.M., Göringer, H.U. (2013). RNA helicases involved in U insertion/deletion-type RNA editing. Biochim. Biophys. Acta. 1829, 835-841.

Künne, T., Swarts, D.C., Brouns, S.J. (2014). Planting the seed: target recognition of short guide RNAs. Trends. Microbiol. 22, 74-83.

Leeder, W.M., Voigt, C., Brecht, M., and Göringer, H.U. (2016). The RNA chaperone activity of the *Trypanosoma brucei* editosome raises the dynamic of bound pre-mRNAs. Sci. Rep. 6, 19309.

Miao, Z., and Westhof, E. (2015). Prediction of nucleic acid binding probability in proteins: a neighboring residue network based score. Nucleic Acids Res. 43, 5340-5351.

Sharma, R., Raduly, Z., Miskei, M., Fuxreiter, M. (2015). Fuzzy complexes: Specific binding without complete folding. FEBS Lett. 589, 2533-2542.

Shoemaker, B.A., Portman, J.J., and Wolynes, P.G. (2000). Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc. Natl. Acad. Sci. USA 97, 8868-8873.

Tompa, P., and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB J. 18, 1169-1175.

Varadi, M., Zsolyomi, F., Guharoy, M., and Tompa, P. (2015). Functional advantages of conserved intrinsic disorder in RNA-binding proteins. PLoS One. 10, e0139731.

Wong, R.G., Kazane, K., Maslov, D.A., Rogers, K., Aphasizhev, R., and Simpson, L. (2015). U-insertion/deletion RNA editing multiprotein complexes and mitochondrial ribosomes in *Leishmania tarentolae* are located in antipodal nodes adjacent to the kinetoplast DNA. Mitochondrion 25, 76-86.

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt, den …….....

.....

Lebenslauf

Persönliche Daten

Name:	Christin Voigt
Geburtstag:	08.04.1988
Geburtsort:	Darmstadt

Ausbildung

09/2012 - 01/2017	Dissertation, Technische Universität Darmstadt
10/2007 - 08/2012	Diplom - Biologie, Technische Universität Darmstadt
09/2004 - 08/2007	Abitur, Gymnasium Michelstadt
09/1998 - 08/2004	Sekundarstufe I, Carl-Weyprecht-Schule, Bad König